

FT	REPEAT	2778	2781	6.
FT	REPEAT	2782	2785	7.
SEQ	SEQUENCE	2867	AA;	331433 MW; 6E7D8CA71AFBFED3 CRC64;
QY	Query Match	4.3%;	Score 231;	DB 1; Length 2867;
	Best Local Similarity	18.94;	Pred. No. 0.0055;	
	Matches 243;	Conservative 219;	Mismatches 436;	Indels 388; Gaps 65;
QY	11	LALTTVSVTYSGEYGLERESVKQEQTQS----	ASEDDWFEEEDNERTKNVSKENSTV-	65
Db	1314	IATTKGTSDTSGDINEL--ESIKEEVHKNLQLVKQESNMEE--	WRKQILISMKDLLIL	1368
QY	66	--DETVDLPSDG-----NSNSSSKTESVVS-----	DPKQVPKAKEPVTQEASN	108
Db	1369	NNSETIAKEISNNTQNALGFRENAKTKLNKTIDELQVAAAMIEAKAHKNNIDIALEDAQ		1428
QY	109	SSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP	PS	168
Db	1429	IDTEVSKIEQINREIMNKKDEIKSYLSEIKERYKCTTEISNSKRGKDKIEPLE-----		1482
QY	169	HAADGTQLTQVASFAPTPDK-----KTAIAEYTSRLGENGKPSRLDID--	QKEIIDEGEI	221
Db	1483	-----KRPNEESNSNKNVINEINENI--RNSQYLLKDIEDAEKQASTKVEL		1527
QY	222	FNAYQLTKLTIPINGYKISIGQDAFVDNKNIAEAVNPESLETISDYAFAMSLKQVLPDNL		281
Db	1528	FHKHE---TTISNIFKE--SEILGVETKSQCKINKAKEDIMKEIERHNSIEIQTVKGFQENTL		1583
QY	282	KVIGELAFEDNQIGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGASQD--	NN	339
Db	1584	NKLNPHNYDN-----AEDELNNDK-----STNAKVLIETNLESVKHN		1621
QY	340	LRNVN-LPDLGLEKIES-----EAFGNPGDE-----	HYNNQVLRTRT	376
Db	1622	ISEITNIKQGGEKIYSKADIMOKIKATSSENTAEKTLKVKYDDQSNVYVNLQIT--	T	1677
QY	377	GQNPQOLATE-----NTVNPDKSL--WRATPMDYTKWLEE-----	DFTYQK	417
Db	1678	ERN--LIVTEKNRLNGIDSTITNEGALKESKGNYEIGFLEKLEIGKNRKLKYDITKKS		1735
QY	418	-NSVTG----FSNKGIOKVRRNKNL-----EIPQKHNGITITEIGDNAFRNV----		459
Db	1736	INSTGVNFSLSLFNNFDLNOYDENKNINDYENKMGELIYNEPEG--SLNKISEN--LRNASENT		1793
QY	460	-DFQS-KTLR-----KYVDL-----EIKLPSTIRKIGAFQSNNLKSFEASEDLEEI-		505
Db	1794	SDYNSAKTLRLLEAQKEKAVNLNKEEANKYLIRDVKVESFRF-----	IFNMKESLDKIN	1847
QY	506	----KEGAFMNNRIGTL-DLKDKLIKIGDAFHINHIAVLVLP-----ESVQEIGRSAFR		555
Db	1848	EMIKKEQTLVNEGHGNVQJVENIKELVDE---NNLSDLIKQATGKNEEIQKITHSTLK		1903
QY	556	QNGALHLMFIGNKVKKTIG---EMAFLSNKLSEVNLSEQOKL-----	TIEVQAFSD	603
Db	1904	NKAKTILGHVDTSAKYVGIKITPBLA-LTELLGDAKLKTQAQELKFESKNNVLETENMSK		1962
QY	604	NALSEVVLPPNLQ-----TIREBAFK-----RNHL		628
Db	1963	NT-NELDVHKNIQDAYKVALLEILAHSEIDTKQKDSKLIENGNOIYLKVVLINQYKNI		2021
QY	629	KEVKG-----STLSQITF-----NAFDQ-----	N	648
Db	2022	STISKEEAVSVKIGNVSKHSHLSKITCSDSKYDNIIALQKOTELQNLRNSTFOEKTNT		2081
QY	649	DGDKRFKG-----KVVVRT-----HNNSHM-----	LADGERFIID	678
Db	2082	NSPSKLEKIKTDFESLKNALKTLEGEVNALKASSDNHEHVQSKSEPVNPALSEIKEETD		2141
QY	679	PKLSTMWDLKVLKIEGLDYSTLRQTTQTOFREMTTA-----GKALLS--	KSNLR	729
Db	2142	IDSILNTALDELKKGRTECVSRKYLKDTVTKEISDDTELINTIEKNVKAYLAYIKQNYE		2201
QY	730	QGEKQFLQEAQFFLGRV-----DLDKAIAKAEKALVTKATKNGHLLERSINKAVLAY		783

Db	NUM1_YEAST	STANDARD;	PRT;	2748 AA.
Db	2202 DTQDVLTLINEHENTKQVSNHEPTNFDKSNKSSER--LTQAVTDSKTIISKL--KGVYIE			2257
QY	784 NNSAIKKANVKRLEREKLDLLTDLVEGKGPLAQATWQGVYLLKTPLPPEYIIGLVNFYD			843
Db	2258 VNENTEMNTIESSAKIEALYNELKNK-----KTSLENIYQTSNEVKLOEMKSNADKYID			2312
QY	844 KSGKLIVYALDMSDTIGEGOKDAYGNPILNVDENEGYHTLAVATLADYEGLYIKDLNNS			903
Db	2313 VS-----KIPFTVLDTQK--SNIVTNOHSINN-----VKDKLKGG			2345
QY	904 LDK-IAKIRQIPLAKYHRLG-IFQAIRNAAAEADRLLEKTPKGYLNEVPNRYKKOMEKTL			961
Db	2346 LQELIDADSSFTLESIKKENEIYSHIKTNIGELEQ-LQQTNKSEHDNVAKHKEKIVH---			2401
QY	962 KPVDYKTPIFNKALPNEKYDGDRAAKGH-----NINAETNNSVAVTPIRSE			1007
Db	2402 -----LINRV---ESLKGD--VKNHDDDQYMKKLNASLINDNIKNTNSINISDEELK			2448
QY	1008 QQLHKSQSDVNLPTQTSKKNFEIYEIL 1033			
Db	2449 KLLKKEVENDQLCKNNNTQNFISDIM 2474			
RESULT 2				
ID	NUM1_YEAST	STANDARD;	PRT;	2748 AA.
AC	Q00402; Q03767;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	10-OCT-2003 (Rel. 42, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Nuclear migration protein NUM1.			
GN	NUM1 OR YDR150W OR YD8358.06.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=5288C / FY1678;			
RX	MEDLINE=97313263; PubMed=9169867;			
RA	Jacq C., Alt-Moerbe J., Andre B., Arnold W., Bahr A., Ballesta J.P.G.,			
RA	Bargues M., Baron L., Becker A., Biteau N., Bloeker H., Blugeon C.,			
RA	Baskovic J., Brandt P., Brueckner M., Butrago M.J., Coster F.,			
RA	Delaveau T., del Rey F., Dujon B., Eide L.G., Garcia-Cantalejo J.M.,			
RA	Goffeau A., Gomez-Peris A., Granotier C., Hanemann V., Hankeln T.,			
RA	Hoheisel J.D., Jaeger W., Jimenez A., Joniaux J.-L., Kraemer C.,			
RA	Kuester H., Laamanen P., Legros Y., Louis E.J., Moeller-Rieker S.,			
RA	Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Paricio N.,			
RA	Paulin L., Perea J., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,			
RA	Prydz H., Punelle B., Rasmussen S.W., Remacha M., Revuelta J.L.,			
RA	Rieger M., Salom D., Saluz H.P., Saiz J.E., Saren A.-M., Schaefer M.,			
RA	Scharfe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,			
RA	Urrestarazu L.A., Verhasselt P., Vissers S., Voet M., Volckaert G.,			
RA	Wagner G., Wambutt R., Wedler E., Wedler H., Woelfl S., Harris D.E.,			
RA	Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentles S.,			
RA	Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,			
RA	Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,			
RA	Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,			
RA	Araujo R., Aviles E., Beino A., Carpenter J., Chen E., Cherry J.M.,			
RA	Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,			
RA	lashkari D., Lew H., Lin D., Moesdale D., Nakahara K., Namath A.,			
RA	Oefner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,			
RA	Shogren T., Shoff N., Winant A., Yelton M.A., Botstein D.,			

OY 991 I 991
Db 1231 V 1231

RESULT 3

```
ALM1_SCHPO STANDARD; PRT; 1727 AA.
ID ALM1_SCHPO
AC Q9UTK5; O13313; Q9UTK8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Abnormal long morphology protein 1 (Sp8).
GN ALM1 OR SPAC1486.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 495-1727 FROM N.A., AND CHARACTERIZATION.
RC STRAIN=972;
RX MEDLINE=20123449; PubMed=10660053;
RA Jimenez M., Petit T., Gancedo C., Goday C.;
RT "The alm1+ gene from Schizosaccharomyces pombe encodes a coiled-coil
RT protein that associates with the medial region during mitosis.";
RL Mol. Gen. Genet. 262:921-930(2000).
RN [3]
RP SEQUENCE OF 644-834 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA Hiraoka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL Genes Cells 5:169-190(2000).
CC -!- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
CC CYTOKINESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
DR EMBL; AL133357; CAB62414.1; -
DR EMBL; AF010473; AAB65416.1; ALT_INIT.
DR EMBL; AB028012; BAA87316.1; -.
DR PIR; T50073; T50073.
DR GeneDB_Spombe; SPAC1486.04c; -.
KW Coiled coil.
FT DOMAIN 57 361 COILED COIL (POTENTIAL).
FT FT 443 463 COILED COIL (POTENTIAL).
FT FT 542 740 COILED COIL (POTENTIAL).
FT FT 804 1106 COILED COIL (POTENTIAL).
FT FT 1223 1427 COILED COIL (POTENTIAL).
FT FT 1497 1555 COILED COIL (POTENTIAL).
FT FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
SQ SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
```

Query Match 4.0%; Score 215.5; DB 1; Length 1727;
Best Local Similarity 19.3%; Pred. No. 0.015;
Matches 231; Conservative 198; Mismatches 422; Indels 343; Gaps 57;

```
OY 12 ALTVSVTVTYQEVYGLERESEVQEQTSASED---DWFEEDNERKTVNSKENVTD 66
Db 116 ALTT-----ENQSLRANSELQEQSKIASQSLAKQDTEALQENSHLGEOVQSAH 167
OY 67 ETVSD-----LFSDGNSNNSSSKTESVSDPKQVPKAKPEVTQEAASNSNDASKVEV 118
Db 168 QALSDIERKKQHMFA---SSSRVKEEILVQE-----KALVSDIASLQSDHSKV-- 215
OY 119 PKQDTASKKETLETSTWEAKDFTVGTDLVGFSGSG--INKLSQTSHTLVPSHAADG--- 173
Db 216 -----CEKLEVVSRQVQDELEK--LAGLAQONTLENEKIQLFEQKRSNYSDDGNTS 264
OY 174 -----TQLTVASFAFPDPKKTALAEYTSRLGENGKPSRLDID---QKEITDEGRIF 222
Db 265 KILETDPSTIKELBEVEVEQKRLTALWE-----SKSSELQSEVALQEKLTSCQSLY 316
OY 223 NAVQLTK-----LTTPNGYKSGIG--DAFVNDKNTAEVN-----LPE 257
Db 317 N-NVTEELNNKQQLISNSLRELQEKYDSVSELQVYKKNKTSVSAGVGLFSPILAQ 374
OY 258 SLETIS--DYAFAMHSLKQVLPDNL-----KVIGELAFPDNOIGKLYLPHRLIKL 307
Db 375 KLSAVQNPFFSFTKVSNNMKLQOKVSSKLQLDLRLTKRFSSFCQV--KQRIIP--VYKQ 430
OY 308 AERAFKSNRIQTVEFLGSKLVYGEASFQDNNLRNVMLPDGLEKIESEAFATGNPDGHHYN 367
Db 431 QSEIVRNRIY-KNFLSESLET-----SNNLTKVQAEILLSTKMRQEAAC-----YL 475
OY 368 NGVVLRTRTGQNPHQ---LATENTYVNPDKSLWRATPDMDYTKMLEDFTYQKNSVTG-- 422
Db 476 QLTASRTQCSDSLREVTICMAELDLNETKS--RNVPAIVQVA-LDE---YAQNPSTASE 529
OY 423 -FSNKGLO-----KVRNRKNLEIPKOHNGITITEIGDNAFRNVDFQSKLTRKYDLEIKL 476
Db 530 TLVNKELANFSSSIKEAVSKTLELRKVRALC-----DVEIQKQTVQYQISNAVKH 580
OY 477 PSTIRKIGAFQSNLKSFEASEDLLEIKEGAFMNR-----IG 516
Db 581 NSNT-----LSEQIKNLESELNSSKIKNESLNERNLKEMLATSRSSITSHNSSAG 632
OY 517 TLDLKDKLIKIGDAFHNHIVAVLPE--SVQETGRSAFRONGALHLMFTGNKYKTIGE 574
Db 633 NID--DKMKSIDESTRELKNEYVRNEMTAIQE---SLSKRNQDL-----LSE 676
OY 575 MAFLSNKLSESVNLSEQOLKT-----IEVQAFSDNALSEVVLPPNLOTI--REAFKFR 625
Db 677 MEAIRKELE--NSKYQOQLSTDRLLTNANDVFAFKEAKEKELRSINQNLQDITISROQRAS 734
```



```

QY 626 NHTKEVSGSSTLSQITFNAFDQNDGDKRFGKVVVTRTHNNSHMLADGERFIIDPKLSST 685
Db 735 KFAEBELHVNSLAERLKGELNASKGEXDLRKRT-----QERLISENDKL--- 778
QY 686 MVDLEKVLKIEGLDYSTRQTOTQTFREMTTAGKALLSKSNLROGEKQFLQEAQFFLG 745
Db 779 LAERERLMSLVSDL-----QTFPLNQ--QQLSDAAR-----KVFSESEKE--- 815
QY 746 RVDLDKAIKAEKALVTKATKNGHLERSINKAVLAYNN-----SAIKKANVRL 796
Db 816 --SLSLQKLKES--NEKMSNDLHSLQKLEKSGIEYSSRIKTLMLKQSLSEDNRLKL 871
QY 797 E-----KELDLITDVEGKPLAQAATWQ-----GVYLLKTPPL-PLPEYITGL 838
Db 872 DNOQMMETIKLQELNGVIELEKQRFSTLEAKFTQOKNTSYSEREALLSSLSLQSKHTSL 931
QY 839 NVYFDSKGLIYALDMSDTIGEGQ---KDAYGNPILNVDENEGYHTLAVATLADYEG 894
Db 932 ESQVNVSLNIEBOLQASKLAEMVERKTEYDERLQTSSELEKNH-LKITSLEQRIYI 990
QY 895 YIKDLINSLDKIKAIRQIPLAKYHRLGIFAIRNMAAHEAD---RLPKTPKGYLNEVPN 951
Db 991 LODEIASSSL-----RCENITKQSETRVALLLEENKHLNNESS 1029
QY 952 YRKQMEKNLKPVDYKTP-----FNRKALPN 977
Db 1030 HENAEKQHLKENDYKQQLLVTEDLKRTREDEYKELLRHADARSTLQKLRDYTKALEQ 1089
QY 978 -EKVDGDRRAK-GHNINAEITNSVAVTPIRSEQQLHKSQSDVNL.POTSSKN-NF 1028
Db 1090 VEDLNKEIALKAGINESQPFPISEKEDPLRQEVVYLKKQNAMLLTLQSSNLF 1143

```

RESULT 4

```

PCPI_SCHPO
ID PCPI_SCHPO STANDARD; PRT; 1208 AA.

```

```

AC Q92351;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Spindle pole body protein pcpi.
GN PCPI OR SPAC6G9.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=21852775; PubMed=11864908;
RA Flory M.R., Morpheu M., Joseph J.D., Means A.R., Davis T.N.;
RT "Pcpi, a Spcl10p-related calmodulin target at the centrosome of the
RL fission yeast Schizosaccharomyces pombe."
RN [2]
RP CELL Growth Differ. 13:47-58(2002).

```

```

RN RP SEQUENCE FROM N.A.

```

```

RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

```

```

RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nure P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -!- FUNCTION: Spindle pole body component that binds calmodulin.
CC Overexpression of pcpi causes the formation of supernumerary SPB-
CC like structures and disrupts both mitotic spindle assembly and
CC chromosome segregation.
CC -!- SUBCELLULAR LOCATION: Spindle pole body.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z81317; CAB03608.1; -.
DR EMBL; AF348506; AAK31344.1; -.
DR PIR; T39068; T39068.
DR GenedB; SPAC6G9.06C; -.
KW Calmodulin-binding; Coiled coil.
FT DOMAIN 151 375 COILED COIL (POTENTIAL).
FT DOMAIN 387 803 COILED COIL (POTENTIAL).
FT DOMAIN 874 1091 COILED COIL (POTENTIAL).
FT DOMAIN 1177 1204 COILED COIL (POTENTIAL).
SQ SEQUENCE 1208 AA; 140763 MW; 70264159AD42424 CRC64;

```

```

Query Match 3.9%; Score 209; DB 1; Length 1208;
Best Local Similarity 19.4%; Pred. No. 0.018;
Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;

```

```

QY 20 TYSQEVYGLERESVKEQQTQSA-----SEDDMEEDNEKNTNSKENVDETVDLS 74
Db 93 SEDPNSYGL---SAISKQATQELSLISQGNDSY---DVKSLTDLK-NSRIDHTDGL-- 143
QY 75 DGNNSNNGSSKTESVSD-----PKQVPKA--KPEVTQEASN 108
Db 144 PANALTLTREQEYLEKVSRENFGRLIKYCLEKRLKLESMAPQIKKAYDNVEIHAERAN 203
QY 109 SSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQ-----T 161
Db 204 LQQLKRTESLQKSEDKNFLE---EKVDYLSK-----VADVEQSQNVKVF 248
QY 162 SHLVLPASHADGTQTLQVA-SFAFTPKTAIAEY-----TSRLGNGKPSRLDIDQK 213
Db 249 ERIRFLENALKEVQREKDSLSTMEEDKSNKEVDYEYRQLQNRRLDEISE--ELDVAQ- 305
QY 214 EIIDEGEIFNAVQTLKLTIPNGYKSGIGDAFVDNKNIAEVLNLP----- 257
Db 306 DLTTEKE---DEIATLKROIKEKENSSSAFENENSSYVHLQEDVAILQAKCDEFADRI 361
QY 258 -----SLRTISDYAFAMSLKQVKLPNMLKYTGELAFDNLQIGKLYLPRHLIKLAERAF 312
Db 362 QVLTADLEKEKENQIMHESEASIGLDSMVY-----H-TLQEQHL 400
QY 313 KSNRIQVFEFLGSKLVIGE--ASFQDNULRNVMPLDGLKIESEAFNGDEHYNNQV 370
Db 401 KAN-EEIEFLHDQISRMNEEGKNFEDIMLQFRSLBEERDVLESKLTQ---LEDNNSL 454
QY 371 VLRTRT-GQNPQOLATENTYVNPDKSLMRATPDMDTKMLBEDFTYQKNVYTGFSNKGLO 429
Db 455 RLMTSSLGNGQIESLRTQNRIDEERKHLRLASKNSDKALAE-----TNIRLQ 502

```

```

QY 430 KVRNRNKLKIPKQNGITTEIGDNAFRN-----VDFQSKTLRKDYDE 472
DB 503 EV--TKELETLRMKNSNDLNEIHDLRRENGGLTKIDSITKEKDRLENEQRIKSYEVN 560
QY 473 EIKLPSTIRK-----IGAFAPQSNLKSFEASEDLLEIKGAFMNRIGT 517
DB 561 VSELNGTIDEYRNKCLKOKEETYNENVMNAFOYKNDLRFHESINKLQDREKELTSN----- 616
QY 518 LDLKDKLIKIGDAAFHINHIAIVLPESVOIGRSAPRONGALHLMFTGNKVKTIGEMAF 577
DB 617 LEKKNLVI-----SSIRETVAMLEKERESIKKYLISGN----- 648
QY 578 LSNKLESVNLSF-----QKQKLT--EVQAFSDNALSEVVLPPNLQITREBAFKRN 626
DB 649 -AKOLDNTNLMELINDKISVLRQRLTDYKDELVDSEEREBAIVAGQKLSASFELMSNEK 707
QY 627 HLKEVKGSSITLSQITFNAFDQNDGDKR-----FGKKV-----VRTHNNSHMLADGERFI 676
DB 708 QALELKYSLSKKNEL-INA--QNLDRREBELSELSKLFEERKIRSGSNDIEKNKEINV 764
QY 677 ID---PDKLS-----STMVDLEKV-----LKTIEGLDYSTLR- 705
DB 765 LNSLADKLQAQIRHLESDDKMLDLVHNLNGIEANIEENAVKRLCLLMGCDYSSVSI 824
QY 706 -----QTTQTOQFREM-----TTAGKALISK 725
DB 825 LQIVSQIEHFVNOQIQTIRSLKQELRHDFVQFSGKKEQELSRSEFEKGLGTETKHDILAQ 884
QY 726 SNLRQGEKQKFLQE-AQFPLGRVD-----LDKALAKAEKALVTKKATK 767
DB 885 RNRNVSEKNDLENAAQKFSSPDRKNGLYPSEHTSKIETYLEKTEIDLKAL--QDELK 942
QY 768 NGHLLERSINKAVLAYNNSAIK-KANVRLKEKELDLTD-----LVEGKG 811
DB 943 NRNLIMDDIS-----SYNKQTKLQEKIKWLERERSILIDELBSYHSNQPNYQNNLVQDKN 998
QY 812 PLAQATMVGVYLLKTPPLPEYIIGLVY--FDKSGKLIYA--LMSDTIGEGQKDAY 866
DB 999 ELEER-----LKEIQKELEYVNNHFMQAEIMTMSVTVDESQMLKTLREAL 1044
QY 867 GNPILNVDENEGHTLAVATLADYEGLYIKDIINSSLDKIKAIQIPLAKYHRL 921
DB 1045 QSKTNNDH-----LSTILERNRKEYKSL-----LDDYNQLR--ARYKNL 1082

RESULT 5
MLP1_YEAST STANDARD; PRT; 1875 AA.
ID MLP1_YEAST
AC 002455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three
RT new open reading frames.";

```

```

RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: Myosin-like protein that is probably involved in DNA
CC repair.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: Ref.2 misquotes the gene name as "MPL1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L01992; AAA34783.1; -.
DR EMBL; X73541; CAA51948.1; -.
DR EMBL; Z28320; CAA82174.1; -.
DR PIR; S38173; S38173.
DR Germonline; 140074; -.
DR SGD; S0001803; MLP1.
DR GO; GO:0005635; C:nuclear membrane; IDA.
DR GO; GO:0005654; C:nucleoplasm; IDA.
DR GO; GO:0006606; P:protein-nucleus import; IDA.
KW Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF. 1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 3.9%; Score 209; DB 1; Length 1875;
Best Local Similarity 19.2%; Pred. No. 0.033;
Matches 221; Conservative 194; Mismatches 375; Indels 362; Gaps 55;

QY 32 ESYVQEQTQASAE-----DDWFEDNERKTNVSKNSTVDETVSGL--FSDGNSN 79
DB 107 ESYVENKMKLSSELEFVKRKLDLDEKKEQTQSNQQRILKILDERLKEIELVVRVNNRSN 166
QY 80 NSSGKTESVSDPKQVPKAKEVTOEASNSNDA-SKYVEPKQDTASKKETT--ETSTWE 136
DB 167 SECKLRSTIMDL-----TKQGYITNDLSRTELERK--TQELTLQSNNDWL 214
QY 137 AKDFYTRGDTLVGFS-----KSGINKL-----SQTSHLVLP SHAADGTQLTQV- 179
DB 215 EKELRSKNEQVLSYRQKTDKVLDIRNELNRLNDFQWERTNNDVYLKQKNNELSKSLQEK 274
QY 180 -----ASFAFTPDKK-----TATAEYTSRLGENGKPSRLDIDOK 213
DB 275 LLEIKGLSDLSNSEKQEFSAEMSLKQRLVDLLESQLNVAVEELNSIRELTAKAVIADDSK 334
QY 214 EIIDGE-IFNAVQTKLTIPNGYK-----SIGQDAFVNDKNIAEVLNLPESLETISDYA 266
DB 335 KQTPENEDLLKELQLTKEKLAQCEKECLRLSSITDEADENENL-----SAKSSSDP- 386
QY 267 FAHMSLKQVKLPDNLKVIIGELAFPDNQIGKLYLPHILIKLAERAFK--SNRIQT-VEF 322
DB 387 -----IFLKQQLIK--ERRTKEHLQNOIETFIYE 413
QY 323 LGSRLKVIGEASFQDNNLRNVMLPDGLEKIESEAFNGPGEHYNNQ--VLRTRTQGNP 380
DB 414 LEHKVPIIN--SPKERT-----DMLENELNNAALLIEHTSNEKNKAVKELNANQ 461
QY 381 HQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSYTGFSNKGLOK--VRNRKULE 438
DB 462 KLVEGENDLQTLTKQ-----RDLICRQIQ--YLLITNSVNSDSKGPLRKEBIOFIQNT- 512
QY 439 IPKQNGITTEIGDNAF--RNVDFQSKTLRKDYDELEIKLPSTIRKIGAFAPQSNLKS 495
DB 513 --MQEDSDTITESDSQKVTERLVEF--KNIIQLQEKNAEHLKVVRNL-ADKLESKEKKS 567
QY 496 FBA-----SEDLBEIKGAFMNNRIGTLDKDLIKIGDAAFHINHIAIVLPESVOEIG 550
DB 568 KQSLQKIESETVNEAKE-AITLKSEKMDLESRIE-----LQKELEEL- 610

```

QY 551 RSAPRONGALHLMFIGNKVKITIGENAFSLNKLSEVNLSEQKQIKTIEVOAFSDNALSEVV 610
Db 611 -----KTSVPNEDASYSN-VTIKQLTETKRDLESQVQ----- 641
QY 611 LPPNLOITREEAFKRNHLKEVSGSTLSQITFNAPDQNDGKRGKVVVRTHNNSHMLA 670
Db 642 ---DIQT-----RISQITRESTENMSLNLKEIQDLVDSKSDISIKLGKESSRILA 689
QY 671 DGERPIIDPDKLSTWVLDLKVLIIEGLDYSTLRQTQTQ--FREMTTAKALLS----- 724
Db 690 E-ERFKLLSNLTLDTKAENDQLRKRPDYLONTILKQDSKTHELNSEVSC-KSKLSIVET 747
QY 725 -----KSNLR-----QGEKQFLQEAQFLGR 746
Db 748 ELNLKKEOKLRVHLEKNLQELNLSPEKDSLRIWVTOLOTLQKREDDLBEET-----R 802
QY 747 VDLDKALAKAEKAL-VTKKAT--KNGHL--LERSINKAVLANNNSAIKKANVKRLKEELD 801
Db 803 KSCQKIDLELDALSELKKEKTSQKHIIKQLEEDNNSNIWYON-----KIEALKQDYE 856
QY 802 LITDLVEGKGPLAQATWGVYLLIKTLPPLPEYIIGLVNVEFDKSGKLIYALDMSDTIGEG 861
Db 857 SVITSVDSK---QTDIEKLQYKYS-----LEK----- 881
QY 862 QKDAYGNPILNVEDNEGHTLAV--ATLADYEGLYIKDIILNSLDKIK-----AIRQIP 914
Db 882 -----EIBEDKIRLHTYVWDETIND-----DSLRLKELEKSKINLTDAYSQIK 924
QY 915 LAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLVNPNYRKQMEKNLKPVDYKTPIFN-- 972
Db 925 EYK-----DLVETTSQSLQGTNSKLDESFKDFTNQIKNLTDKETSLEDKISLKEQMFNLN 980
QY 973 -----KALPNEKVDGDRAKGNHINAEFTNNSVAVTPIRS--EQQLHKSQSDVN--LP 1020
Db 981 NEILDQKKEKEKADPKK-----RISILQNNKEVEAVKSEYESKLSKIQNDLDQGTIY 1035
QY 1021 QTSKKNPIYEI 1032
Db 1036 ANTAQNNYEOEL 1047

RESULT 6
SCA4_RICRH STANDARD; PRT; 1013 AA.
ID SCA4_RICRH
AC Q9AJ81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia rhidicephali.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=33992;
RN [1]
RE SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene D' coding for an intracytoplasmic protein";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155053; AAK30684.1; -.

KM Antigen.
FT NON TER 1 1
FT NON TER 1013 1013
SQ SEQUENCE 1013 AA; 110550 MW; 856E98912315D102 CRC64;
Query Match 3.9%; Score 208; DB 1; length 1013;
Best Local Similarity 19.4%; Pred. No. 0.016;
Matches 194; Conservative 148; Mismatches 322; Indels 336; Gaps 47;
QY 53 ERKTVSKENSTVDETVSDLSFDGNSNNS-----SSKTESV-VSDPKQV----- 95
Db 108 EOKRKEIEEKEKDKTLSTFF--GNPANREFIDKALENDELKKELESIEIAGYKVVHNTF 165
QY 96 -----PKAPPEVTQEAASSNDASKVEVPKQDTASKETLETSTWEAKDF-VTRGD-TL 147
Db 166 STASGYPGGFKPVQWENQVSASDL-RATVKNADAGDELCTINETVTKTPEFTVAKQDGTQ 224
QY 148 VGEFSK-----SGINKLSQTSHLVLPSPAADGTQTLQVASFAPTPDKTAY--AEYTS 197
Db 225 VOISSYREIDFPKLDKADGSMHLSMVALKADGTK-----PSKDKAIYFTAHYEE 274
QY 198 RLGENGKPSRLDIDQKEIIDEGEIENAYQTLKLTIPNGYKSIQDAFVDNKNIAEVNLE 257
Db 275 --GPNKGP----- 280
QY 258 SLETISDYAFAMHSLKQVLPDNLKVIG---ELAFEDNQIGKLY--LPRHLIKLAER 310
Db 281 -----QLKEISPKPLKFGAGTGDAIAYIEH--GGEITYTLAVTRGKYKEMMK 325
QY 311 AFKSNRIQTVFPLGSKLVIGEASFQDNNL-----RNMLPDGLEKIESEAFGTGNG 362
Db 326 EVELNQGSVDLSQAEIDITIGQGSKEQPLITPQQTASSSVESFYKQQVPPITPTNQ- 384
QY 363 DEHYNNQVLRTRTQGNPHQATENTYVNPDKSLWRATPMDYTKWLEEDFTYQKNSVTG 422
Db 385 -----LQPETSQMPQSQQVNPMLNNAATALSGMQDL-----LN 418
QY 423 FSNKGLQK-VRRNKNLEIPKO-----HNG-----ITTEIGDNAPRN----- 458
Db 419 YVNAGLTKEIDSNKQIDILKEAATAILHNAKSDIAEKQNTITALAENTVNNQNLTPDAKV 478
QY 459 --VDFQSKTLRK---YDLEERK-LPSTIRKIGAFAPFQSNLKSFEASEDLSEIKEGAFM 511
Db 479 AGVNAVLETIKNDQNTPLEKSKMLEATV---AITNSENLPEPKQKEQMLEKTVD--- 530
QY 512 NNRIGTLDLKDKLIKIGDAFHINHYAIVLPESVQEIGRSAFRONGALHLMFIGNKYKT 571
Db 531 ---VG-LSLKDD---ASRAAAIDGITDAVIKSNLS-----TEDKGTMLIAGVDKY-- 573
QY 572 IGEMAFSLNKLSEVNLSEQKQI--KIEVQAFSDNALSEVVLPPNLOITREEAFKRN 626
Db 574 -----NVELSNAEKQKILGSLKKGVEAQVLS--PAQQQLMQQNLDKITAEQTKD 623
QY 627 HIKEVKG-----SST-----LSQITFNAPD-----ON 648
Db 624 TIKKVNLDLFDPLSLSTELKTNNIQAITSNVLDPATAVYKGEIIQETNTVAGSSIEAQD 683
QY 649 DGDKRFGKVVVRTH-----NSHMLADGERFIID-----PDK--LSSTMVD----- 688
Db 684 KAEIVKGVETIATHSDTSLSPNKALIMASAEKGVESKTNLPDRELMTKGLVDGIYEG 743
QY 689 ---LEKVLKIEGLDYSTLRQTQTQFREMNTTAKAL--LSKSNLRQGEKQKQLOE--- 739
Db 744 KCGPEITKAVSSGIDNSINNSEKEALKKAKADAASEATLDIETQNLTEGLKQGNIEHKP 803
QY 740 -----AQFLGRVD-----LDK--AIKAKEKALVTKKATKNGHLLEERSINK-----A 779
Db 804 RDDIYNKAQEVINAVNPVLEALEKPAVSAEERIYQETSSILNNISKLAVEKANNFR 863
QY 780 VLAYNNSAIKKANVKRLKEELDILTDLVEGKGPLAQATWGVYLLIKTLPPLPEYIIGLN 839
Db 864 MLSSNGNF--KTLKKKEESIKKVDLVKAFG-----TKSSTEQGSFTKAN 908

QY 840 VYFDKSGKLIVYALDMSPTI---GEGQDAYGNPILANDE 875
Db 909 LIDDKTLISKEVRLQITIDKLOEQOTQQAIAIENPSVKTED 948

RESULT 7
SCA4_RICAK
ID SCA4_RICAK STANDARD; PRT; 998 AA.
AC 09AIX9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia akari.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=786;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF213016; AAK30691.1; -.
KW Antigen.
FT NON_TER 1
FT 998 998
SQ SEQUENCE 998 AA; 109328 MW; FCEB43AC62DE5BD5 CRC64;

Query Match 3.8%; Score 203; DB 1; Length 998;
Best Local Similarity 18.9%; Pred. No. 0.027;
Matches 220; Conservative 176; Mismatches 405; Indels 360; Gaps 55;

QY 6 LKTIALALTYSVTVYQEVYGLERESEVKQEQTSASE-----DWMFEEDNERKT 56
Db 57 ISTLSGTISTDDQI--SDPITKAVREIILIQQKDEIAEQILKDLAALVDRDLAEQKRKEI 114
QY 57 NVSKENSTVDETVSDLFSDGNSNNSSTKTESVSDPKQVPKAKE-----VT 103
Db 115 EEEKEK--DKKLSVFF--GNPAN-----REFIDNALKEPELKKKLESEITGYKNILLT 164
QY 104 QEASN-----SSNDASKVEVPKODTASKEETLETSTWEAKDFVTRGDTLV 148
Db 165 YSAANGYHGFKPYQWENQISASDL-RATVVKNADAGDELCTINETTVKTKPFTVA----- 218
QY 149 GFSKSG-----INKLSQTSILVLPASHADGTQLTQVASFAPTPDKKTAI 192
Db 219 --KKDGTQVQINSYRAIDPEPIKLDKADGSMHLSMVALKADGTK-----PSKORAV 266
QY 193 ---AEYTSRLGNGKPSRLDIDQKEIIDEGEIFNAVQLTKLTIPNGYKISQGDADFVNKN 249
Db 267 YFTAHYEE--GPNKGP----- 280
QY 250 IAEVNLPESETTISDYAFAMSLKQVKLPDNLKVIQE---LAFPDNQIGKLY---LPR 302
Db 281 -----QLKRISSPQPLKFAAGDPDAVAYIEH--GGEITYTLAVTR 317
QY 303 HLIKLAERAFKSNRIQTV---EFLGSKL-KVIGEASFQDNNLRNMLPDGLEK--IESEA 356
Db 318 GKCKEMMEVELHOGQGSYDLISQIIABDLTKVQGRS---QETLQPIITPNQELKSSIEPTPT 374

QY 357 FTGNPGDEHANNQVYLRTRTGQNPQOLATENTYVNPDKSLWRATPDM-----DYTKMLEE 411
Db 375 TTQVPPRITPASQPV--HTETSQMP-----QSQQVNP--NLFNAAATALSCSMQDILLNVNA 425
QY 412 DFTYQK--NSVYGFSPNKGLOKVRNRKNLEIPKOHNGITITE--IGDNAF-----RNVD 461
Db 426 GLTKEKDQNTQIDLINEAATAILNN---EKEQANFTILTNNVNNNALLPTDKVARVNA 482
QY 462 QSKTLR-KYDLBEIKLPSTIRKIGAFAFQSNLKSFEASEDLLEIKE----- 507
Db 483 VLETIKNNQDTPLEKSKMLEATVAITLNSENLTPKQKQOMLEKAVDVLSPKDDTSRAV 542
QY 508 -----GAFMNNRIGTLIDLKDLIKIGDAAFHNIYAILPVESVQEIQRSAFRONGALH 561
Db 543 AIDGITGAVIKSNLSTKDKGTMILAVGDKV-----NASELSNAEKQ----- 584
QY 562 IMFIGNKVKTIGEMAFLSNKLSEVNLSEQKQLTEVQAFSDNALSEVVLPPNIQTIREE 621
Db 585 --LLGSVLKKGVEETKILS-----PEQQQL-----MQQNLDKITAE 617
QY 622 AFKRNHLKEVKGSSITLSQITENAFDQNDGDKRFQKVVVTRHNSHMLAD--GERFIIDP 679
Db 618 QTKNDNITEVQG--ILANDAFNTIAKTAIQKVTIKVL-----DSPITAEIKGE----- 664
QY 680 DKLSSTWVLDLEKVLKITEGIDYSTLRQTTQTOFREMTTAGKALLSKSNLRQGEKQFLQE 739
Db 665 -----TLESITKIYA---ESPLNVQDKTDI--VKMGGEALASHRTMAPTKIAAIE- 710
QY 740 AQPFLLGRVLDKAIKAERKALVTKKATNGHLBERSTINKAVLAYNNSAIKKANVRLKE 799
Db 711 -----SVETGVAKSITDIEDKMLTKGLVDGIYEDKA-----NPEITSEMMKAVSKG 757
QY 800 LDLLTDLVEGKPLAQ-----TWQGV--YLKTPPLPEYITGLNVYFDKSG 846
Db 758 VDNSTAIPEDKQALKDAASEALDRATQNTFELGKGNLDEPKPRDDIY-----NKAQ 810
QY 847 KLIYALDMSDTTIGEGQDAYGNPILNVEDNEGYHTLAVATLADYEGLYIKDILNSSLDK 906
Db 811 DIAYALK-----NVVTTVLDPANPEKREVSSEBEVMNKTSSILNDISKIAIEK 856
QY 907 IKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLNEVPENYRKQ--MEKNLKPVD 965
Db 857 VNNLRAM-LSPDSNL--KTLEEKKAATKQVDELVKREFGTSSTEEQOSFIQANL--ID 910
QY 966 YKT-----PIFNKALPNEKVDGDRRAAGHINAE---TNNSVAVTPI-RSEQQLHKS 1013
Db 911 DKTLSEKVRQLQITDKLQEQAQKRAEAIKPNVKTEDLRAVSGQSALKPISNDEPDIEKT 970
QY 1014 QSDVNLPEQTSKNNFIYEILG 1034
Db 971 KMVVGDRVNIKDN--KIMG 989

RESULT 8
RBP1_PLAVB
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RL merozoites."
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: Involved in reticulocyte adhesion. Specifically binds to

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AJ235272; CA14951.1; ALT_FRAME.
 DR EMBL; AJ235272; CA14950.1; ALT_FRAME.
 DR EMBL; AF200340; AK31305.1; -.
 KW Antigen; Complete proteome.
 FT CONFLICT 11 15 EFDPL -> RPGLV (IN REF. 2).
 FT CONFLICT 365 365 H -> Y (IN REF. 2).
 FT CONFLICT 413 413 MISSING (IN REF. 2).
 FT CONFLICT 957 957 G -> R (IN REF. 2).
 SQ SEQUENCE 1022 AA; 11410 MW; 03230E3A663A9622 CRC64;

Query Match 3.8%; Score 202; DB 1; Length 1022;
 Best Local Similarity 18.8%; Pred. No. 0.031;
 Matches 207; Conservative 196; Mismatches 400; Indels 298; Gaps 54;

QY 28 LERESVKOE---QTQSASEDDWFEEDNERKTVNSKENSTVDETVSDDLPSDGSNNSSSK 84
 DB 85 LEKQRDILREYFVNTNPELAQIAKEEDRKFRFLSNQDNVALINKAFEDTKTKNLEK 144
 QY 85 TE-----SVVSDPKQVPKAPPEVTQZASNSSNDASKVEVPKQDTASKKETLETSTWE 136
 DB 145 AEIVGYKNVLSTYSVANGYGGFQPVQWENQVSASDL-RSTVVKNDSEBELCTLNETTVK 203
 QY 137 AKDFVT--RGDTLVGFSK-----SGINKLSQTSHTLVLPASHADGTQTLQVASFAPTPD 187
 DB 204 TKDLIVAKQDGTQVQINSYREINFPILKDKANGSMHLSMVALKADGTYPKADKAVYFT-- 261
 QY 188 KKTAAIEYTSRLGNGSKPSRLDIDQKEIIDGEIFENAYQLTKLTPNGYSIGQDAFVDN 247
 DB 262 -----AHYBE--GPNKGP----- 272
 QY 248 KNIAEVLNPESLLETISDYAFAMSLKQVKLPDLNKVIG---ELAFPDNQIGKLY---L 300
 DB 273 -----OLKEISSPOLKFKVGTGDVAVAYIEH--GGEITYTLAV 307
 QY 301 PRHLIKLAERAFKSNRIQTV--EFLGSKLVKVGASFDNNLRNVMLPDGLKIESEAF 357
 DB 308 TRGKYKEMKEMKEVALNFGQSVALSQTIAEDLTHTVQSPSHETH--KPIITPN--QELLESI- 362
 QY 358 TGNPGEDEHYNNQV-----VLRTRTGQNPQHLATE-----NTYVNPDKSLMRATP 401
 DB 363 -----EQHTSQQVPPITTFNKSLQPKISQ-IHQLPQQAQSSGIPNPVLANAANALSTSMQ 416
 QY 402 DM--DYTKMLEEDFTYQNSVTGFSNKGLOKVRNRNLEI--PKOHNGITITETGDNAFRN 458
 DB 417 DLNNINSYLTKNQDINKQS--DLIKEAIAIALNNKKSDFAEKQYN--ITDLAKNIFSN 471
 QY 459 VDFQSKTLRKYLEIKLPSTIRKIGAFQSNMILKSFEASEDLKEIKEGAFMNNRIG-- 516
 DB 472 KDIIADA-----KVVNVNTL-----LETIQNDQNTLIDIKSKILEDVAIT 512
 QY 517 -----TLDLKDK--LIKIGDAAFHI--NHIYAVLPESVQEIGRSAFRONGALHLMFIGNK 568
 DB 513 LNSENIELKQKQILKEKVDIGLSIKDDISRVAVADSIMD--TVIKSN-----IANE 562
 QY 569 YKTIGEMAFLS-----NKLSEVNLSEQQL-----KTIEVQAFSDNALSEVVLPLPLOTI 618
 DB 563 DK---EKIFITVFQDQINSYFENAVAKQKLLDSILKKTAEQVLSPE--QQQLMNQNLDMI 617
 QY 619 REEAFKRNHLKEVKG-----SSTLSQITFNAFDQNDGDKRFGKKVVVRTHNNSHMLADG 672
 DB 618 TTEHTKRDITIEKVNMLLEPLSNALTKT-----NIQVMTSN----- 654
 QY 673 ERFTIDPDKLSTMDLEKVLKIEGLDYSLRQTQTOFREMTTAGKALSKSNLROGE 732
 DB 655 ---VLD-----SPVQIEKSKLIQVVT-KTVAESALVEPKDTEIVKG-IGKTIYTHSD 703
 QY 733 KQKFIQEAQFFLGRVLDKAIKAKALVTKATKNG--HLEERSINKAVLAYNNSAIK 789
 DB 704 TSLPLHDKVVMGVS--AKGIVESKNDLLDRELIAGLVGDGYEAKGDNAVVAHAISSMIA 761

QY 790 KANVKRLKE-IDLLTLVEGKGPLAQATMWQGVYLKTPLPPEYIIGLVNYPDKSKL 848
 DB 762 NSNINQSEKALKRSQDVSEK-----VLDKEIQ-----NLDRLEKQON 800
 QY 849 IYALDMSDTIGEGQKD--AYGNPILNVEDNEGYHTLAVTLADYEGLYIKDILNSSLD 905
 DB 801 INESKLHDDIYNKTQDVANALKNVITTVLDDNSGQRGVSEAPKVVSSL-LINDISKRTIE 859
 QY 906 KIKAIRQIPLAKYHRLIGIQAIRNAAEADRLPKTPKGYLN-----EVPNRYKKQMEKN 960
 DB 860 KINMLRAM-LSQDGNLKTPEEKD--EATKKVDELVYAFDNKSSTEQONFIKSNL--- 912
 QY 961 LKPYDYKTPIFNKALPNEKVDG-----DRAKGNHINAETNNSVAVT-----PI-RS 1006
 DB 913 ---IDNKT--LSREIRLQIIDNLLKAQAQKRAETIENLSAKTEDVRVISGKSELKPIQSD 967
 QY 1007 EQQLHKSQSDVNLPTQSSKN 1027
 DB 968 EPIYQAKKVVVERDRVIDKDN 988

RESULT 10

TE95_TETTH STANDARD; PRT; 872 AA.
 ID TE95_TETTH
 AC Q94819;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Telomerase component p95 (EC 2.7.7.-).
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymenina; Tetrahymena.
 OX NCBI_TaxID=5911;

[1]
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RP MEDLINE=95292335; PubMed=7774009;
 RX Collins K.; Kobayashi R.; Greider C.W.;

RA "Purification of Tetrahymena telomerase and cloning of genes encoding
 RT the two protein components of the enzyme.";
 RL Cell 81:677-686 (1995).

CC -1- FUNCTION: Ribonucleoprotein DNA polymerase that catalyzes the de
 novo synthesis of telomeric B1mple sequence repeats. P95 contains
 CC some or all of the template-independent primer DNA-binding site
 CC termed the anchor site.

CC -1- SUBUNIT: Telomerase consist of two subunit, p80 and p95 that form
 CC a 1:1:1 complex with the 159 nt telomerase RNA.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; U25642; AAC46602.1; -.
 DR PIR; S55940; S55940.
 DR Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
 KW DNA-binding.
 KW SEQUENCE 872 AA; 102946 MW; 78AFF4CB829651FA CRC64;

Query Match 3.7%; Score 200; DB 1; Length 872;
 Best Local Similarity 20.3%; Pred. No. 0.031;
 Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTVSVTVYSQEVYGLF-----REESVK-----QEQTSASEDDWFEEDNERK- 55
 DB 17 TNLDFVLQNLLEVYKSQLBHYKTQOQKIEEDLKILKFNQDQDGNSGNDDDENNNSNKQ 76
 QY 56 -----TNVSKENSTYDETVSDDLPSDGSNNSSSKTESVSDPKQVPKAPPEVTQZASN 108

Db	77	QELLRRVNOIKQOVOLIKKVGSKVEKDLNLN-----EDEN	111
QY	109	GSNDASKVEVPKODTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG-----	154
Db	112	KKNGISEQOV-----KEEQLRITTEEQVKKYQNLVENMDYQLDLNESGGHRRHRET DY	164
QY	155	-INKLSQTSHLVLP SHAADGTQLTQVAFPAFTPDKKTALAEYTSRLGENGKPSRLDIDOK	213
Db	165	DTEKWEFEISH-----DQKNVYSIYA---NQKTSYCWMLKDYENKNNDHLNVSIN	211
QY	214	EIIDEGEIF---NAYQLTKLTIPNGYKSIGODAFVDNK-----NIAEV-	253
Db	212	RLETEAEFYAFDDFSQTIKLT-INNSYQTWNIDVNFDDNNLCILALRFLSLBERFNILNIR	270
QY	254	-----NLPSLETISDYAFAPAMSLKQVKLPDNLKVIGELAFPDNQIGK-----	297
Db	271	SSYTRNOYNFEKIGELLETFIAVVFSSHRIQGIHLQVPCEAFOYLVNSSSQISVKDSQLQ	330
QY	298	LY-----LPRHLIKLAERAFKSNRIQTVEFLGSKLVIGEAS	334
Db	331	VYSFSTDLKLVDTNKVQDYEFKPLQOEPR-LTHVSQQAIPVSATNAVENLNVLKKVKAH-	388
QY	335	FQDNNLRNVMLPDGLEKIESEAFITGNPGDEHYNNQVLRTRTGONPHQIATENTYVNPDK	394
Db	389	-----NLNLVSI-----TQFNDFYFVNLQHLKLEFGLBNILTKOKL-----E	428
QY	395	SLMRATPDMDYTKWLEEDF-TY--QKNSVTGFSNKG--LOKVRRNKMLEIPKQHNGITIT	449
Db	429	NLLISIKQSKNLKFLRLNFYTYVAQETSRKQILKQATTIKNLKNNKNOEETPBTKDETPS	488
QY	450	EIGDNAFRNVDFOSKTLRKYDLEIKLPST---IRKIGAFAFQSNMLKSPAS--EDLEE	504
Db	489	E-STSGMKPFHDHSELTELEDF-SVNLQATQEIYDSLHKLLIRSTNLKKEKLSYKEMEK	546
QY	505	IKEGAF--MNNRIGTL-DLKDKLIKIGDAFHINHIIYAVLPESV-----QEI--G	550
Db	547	SKWDTFIDLKNITYETLNMLKRCSVNISNP--HGNISYELTNKDSIFYKFKLTLNQELQHA	604
QY	551	RSAFRQNGALHLMFIGNKVKITIGEMAFSLNKLESV-----NLSEQOLKTIIEVQAFS	602
Db	605	KYTFKQN-----EFQFNNVKS---AKIESSSLESLEDIDSLCKSIASCNULONV-----	650
QY	603	DNALSEVVLPPNLQOTIREAEFKRNHL---KEVKGSSTLSQITENA-FDQNDGDKRFGKKV	658
Db	651	-NIASLLYPNNIQ---KNPFNKPNLLFFKQFQOLKNLENVSINCILDQ-----	695
QY	659	VVRTHNNSHMLADGERFIIDPDKLSTWVDLEKVLKIEGLDYSTLRQTO-----	709
Db	696	-----HIINSISEFLEKNKKIKAFILKRYLLQYY--LDYTKLFPTLQOLPBLNOVY	745
QY	710	--TQREMTAGKALLSKSNSLRQEKOK-----FLQEAQFFLGRYVLDKAIA--KA	756
Db	746	INOOLEELTVSE---VHKQVWENHKQKAFYEPPLCEFIKESSQTLQOLIDPDQNTVSDSI	801
QY	757	EKALVTKATKNGHLLERSINKAVAIYNNSAIKKANVKRLEKELDLITDVEGKGPLAQ	816
Db	802	KKILESISESKYHHYLRINPSQ---SSSLIKSEN---EELQELLKACDEKGVLVKA	851
QY	817	TMVQGVYLLKTPLPPLPEYYIGLVNYFD	843
Db	852	-----YKFPPLCLP-----TGTYYD	866

```

RESULT 11
1F2_WOLSTU
ID_1F2_WOLSTU          STANDARD;          PRT;          939 AA.
AC_Q7M7X5;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR WS2016.
OS Wolinella succinogenes.

```

```

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=DSMZ 1740;
RX MEDLINE=22882897; Pubmed=14500908;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klimmek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
CC -1- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the IF-2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BX571662; CAE11018.1; -.
DR HAMAP; MF_00100; -; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
KW DOMAIN 441 589 G-DOMAIN.
FT NP_BIND 447 454 GTP (BY SIMILARITY).
FT NP_BIND 493 497 GTP (BY SIMILARITY).
FT NP_BIND 547 550 GTP (BY SIMILARITY).
SQ SEQUENCE 939 AA; 102670 MW; B3D7BD9FBECE3BC0 CRC64;

```

	Query Match	3.7%;	Score 200;	DB 1;	Length 939;	
	Best Local Similarity	19.7%;	Pred. No. 0.034;			
	Matches 214;	Conservative 148;	Mismatches 365;	Indels 360;	Gaps 49;	
QY	29 ERESVKQ--EQTQSASEDDWFE-EDNERKTNVSKENSTVDETVSDLFSDGNSNNSSKST	85				
	: :	:	:	:	:	:
Db	92 EVKESVKEAPESLPESPKEEAFEAIPKESVKVTPTKLEQEPPELVSIEPSLESASET	151				
QY	86 ESVSDPKQVPKAKPEVTQEAASNSSNDASKVEVPKODTASKKETLETSTWEAKDFVTRGD	145				
	: :	:	:	:	:	:
Db	152 ---LSDSNPLPQEKTE-TKETIVATTLATQTDAEIQSEBEKETLAQATVQR-----	200				
QY	146 TLVGFSKSGINKLSQTSHLVLPSHADGTQLTVASFAFTPDKKTAIAEYTSRLGENGP	205				
	:	:	:	:	:	:
Db	201 --VGLRI--YKKRSEE-----PAKADRPLEE---ARTPSRTAGLKTLLQSLGES---	244				
QY	206 SRLDIDQKEIIDEGEIFNAVQLTK----LTIPNGYSIGQDAFVDNKNIAEVNL--PES	258				
	:	:	:	:	:	:
Db	245 -----DESEALARKKKKKKKPLPAFT-----KNEQKIDLLGDRA	281				
QY	259 LETISDYAFAHMSLKQVKLPDLNKVIGELAFPNQIGKLYLRHLIKLAERAFKSNRIQ	318				
	:	:	:	:	:	:
Db	282 LETVSSF-----DDEQE--EIVLFDLTI-----RDDINKEDEVAK--KVD	317				
QY	319 TVEFLGSKLKVTIGEASFQDNMLRNVM.LPDGLEKIESEAFTGNPGDEHYNNQVVLRTRTGQ	378				
	: :	:	:	:	:	:
Db	318 T-----DRIKVQRKTPFLDQIRRV-----KRRKR	342				
QY	379 NPHOLATENTYYNPDKSLMRATPDMDYTKWLBEDFTQKNSVTGFSNKGLOKVRNRKNLE	438				
	: :	:	:	:	:	:
Db	343 RPQTVA-----DKESISGTIETP-----EEIRAVEFAEKTGKSIGEVIKVLFNLGLM	389				
QY	439 IPKHNGITITTELGDNAFRNVDFOSKTLRKYLDLEEILPSTIRKIG-AFAFQSNNLKSF	497				
	:	:	:	:	:	:

```
Db 390 ITRN-----DFLDSDSIEILAELELDVVIKNTSEALEYTSER---E 428
Qy 498 ASDELEIKEGAFMNNRIGTLD-----LKDKL-----IKIGDAAFHINHIAVLPESVQE 548
Db 429 EDEDEGLEERPPVVTIMGHVHGKTSLLDKIRNTKVAAGEAGITQHIGAYTVEKDGK 488
Qy 549 IGRSAFRONGALHLMFIGNKVKTIGEMAFLSNKLSESVNLSEKQKLTIEVQAFSDNALSE 608
Db 489 I-----SFIDTPGHEAFTEMRARGAEVTD-----IVIIVLAADGVKQ 526
Qy 609 VVLPENLQITIREAFKRNHLKEVGSSTLSQITFNAPDQNDGDKRFGKVVVRTHNNSHM 668
Db 527 -----QTI--EA-----LNHAKAANVPITIALNKVDKEDA----- 554
Qy 669 LADGERFIIDPKLSTMTVDLEKVLKIEGLDYSTRQTQTFQREMTTAKALLSKNL 728
Db 555 -----NPDVKKAEAD-----LGYSPLWEGEYEFVHIS----- 583
Qy 729 RQGEKQKFLQEAQFFLGRVLDKAIKAERKALVTKKATNGHLLERSINKAVLANNNSAI 788
Db 584 -----AKTGEI-----DHLE-----TILVQSELE 605
Qy 789 KKANVRKLEKELDLTDLVEGKPLAQATWQVYLLKTPPLPEYYIGLVNVPKSGKL 848
Db 606 LKANPEKAKAVVIESLEKGGVATVIVQSGTLKVGDSTIVADTAYGRVRLIDDCGKN 665
Qy 849 IYALDMSD---TIGEGQKDAYGNPILNVDEDENGYHTLAVATLADYGLYI--KDIINS- 902
Db 666 IQSIGSEVAVVTGLSETPMAGAVLSVENDS-----IAREVAEKRALYLRQKELSGST 719
Qy 903 --SLDKIKA-----IRQIP-LAKYHRLGIFQAIRNA-----AAE 933
Db 720 KVSFDELASAVAEQGLKSLPVITIKADTQGLSKAIRGSLKELNBEVKINIHAAGVGITTE 779
Qy 934 ADRLLPKTPKGYL---NEVPNRYKKQMEKVLKPVYKTPITPKALPNKVDGDRACKH 989
Db 780 SDVVLGASDNSVILGFNVPRPTGSYVKNRAKEL-GVEVKTYISIIYAL---LDDVRAVLGG 834
Qy 990 NINAETNSVAVTPIRSEQQLHKSQSDVNLPTQSSKNNFYIYELIGVSLCLFLVTAG-- 1047
Db 835 -----MMSPLVEE-----NTGQAFVRETFTIARVGTIAGC---LVTDGSI 872
Qy 1048 KKKGRAR 1054
Db 873 QRGIXVR 879

RESULT 12
GOBI_HUMAN STANDARD; PRT; 3259 AA.
AC Q14789; Q14398;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
DE (Golgi complex-associated protein, 372-kDa) (GCP372).
GN GOLGB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94187728; PubMed=7511208;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT protein (giantin).";
RL Mol. Cell. Biol. 14:2564-2576 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94257116; PubMed=8198703;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
```

```
RA Renz M.;
RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.";
RL J. Autoimmun. 7:67-91 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Schda M., Misumi Y., Fujiwara T., Nishioke M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RT localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408 (1994).
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the golgin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; X75304; CAA53052.1; -.
CC DR EMBL; D25542; BAA5025.1; -.
CC DR PIR; A56539; A56539.
CC DR PIR; I52300; I52300.
CC DR Genew; HGNC:4429; GOLGB1.
CC DR MIM; 602500; -.
CC DR GO; GO:000139; C:Golgi membrane; TAS.
CC DR GO; GO:005795; C:Golgi stack; TAS.
CC DR GO; GO:0016021; C:integral to membrane; TAS.
CC DR GO; GO:007030; P:Golgi organization and biogenesis; TAS.
CC KW Golgi stack; Antigen; Coiled coil; Transmembrane.
CC FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 3236 3256 POTENTIAL.
CC FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
CC FT DOMAIN 48 593 COILED COIL (POTENTIAL).
CC FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
CC FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
CC FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
CC FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
CC FT DOMAIN 2420 2423 POLY-GLU.
CC FT DOMAIN 2293 2296 POLY-SER.
CC FT CONFLICT 1 39 MISSING (IN REF. 3).
CC FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
CC FT CONFLICT 1765 1765 D -> G (IN REF. 3).
CC FT CONFLICT 2950 2950 H -> D (IN REF. 3).
CC SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178BD CRC64;

Query Match 3.7%; Score 199.5; DB 1; Length 3259;
Best Local Similarity 20.7%; Pred. No. 0.2;
Matches 217; Conservative 168; Mismatches 386; Indels 277; Gaps 52;

Qy 9 LALALTVSVVTVYSEQEVYG-LERRESVKQEQTO-----SASEDDWFEDNERKTNSKEN 62
Db 326 LEVAKRKLSFHNLOEEMHHLLLEQFEQAGQAQAELESRYSALEQKHAEMEKTSHTLSIQ 385
Qy 63 STVDETYS--DLFSDGNSNNSGSKTESVSDPKQVPKAKPEVTQEAASNSNDASKVEVPK 120
Db 386 KTGELOSACDALKDQNSKLLQDKNEQAVQSAQTICQLEPDLQOKSKEISQFLNRLPLQO 445
Qy 121 QDTASKEETLETSTWEAKDFVTGDTLVVGFSGKINKLSQTSHLVLPASHADGTQLTQVA 180
Db 446 HETV-----SQTS---FPDVYNEGTO----- 463
Qy 181 SFAFTPKKTAIAEYTSRLGENGKPSRL--DIDQKEITDEGEIFNAYQLTKLTIPN----- 234
```

Db 464 --AVTEENIASLQKRVVEL-ENEKALLSSIELEELKAENEKISS-QITLLEAQNRTE 519
QY 235 GYKSIGQDAFVNDKNIAEYVNPESLETISDYAFA--HMSLKQVKLPDNLKVIGELAFPDN 292
Db 520 ADREVEISIVDIANKRSSAESGQDLVENTFSQKHKL-SVLLLEMEKAQEELAFKL 578
QY 293 QIGKLYLPRHLIKLAERA-----FKSNRIQTVEFLG---SKLVIGEASFQDNNLRNVL 345
Db 579 QLOG-----KRAEEDHEVLQDKEMKQMEGELAPIKMKVPLEDTGQDFPL----M 625
QY 346 PD---GLEKIESEAFGTGNGDEHYNNQVLRTRTGQ---NPHQLATENTYVNDKSLMR 398
Db 626 PNEESSLPAVEKE---QASTEH-----QSRITSEISLNDAGVELKSTKQDGDKSL-S 673
QY 399 ATPDMYTKWLEEDFTYQKNSVTGFSNKLQKVRNRKNLEIPKQHNGITTEIGDNAFRN 458
Db 674 AVPDIGQCH--QDELERLSQIL-----EEL-NEHKAQEIYE-----KN 710
QY 459 VDFQSKTLRKVD--LEETKLPSITRKIGAFAFQSNMLKSFEA-SED---LEETKEGAFM 511
Db 711 LDEKAKEISNLNQLIEEFKNA-----DNSSAFALSEERDQLSQVXELSMV 759
QY 512 NN---RIGTLDL-----KDKLIKIGDAFHINHYAIVLPESVQEIGRSAFRNGALHL 562
Db 760 TELRAQVKQLEMNTLAERQGRRLDYESQTAHDN-----LLEQI-----HS 800
QY 563 MEIGNKVKTIGEMAFLSNKLSEVNLSEQOKLTI-EVQAFSDNALSEVLPNLTIREE 621
Db 801 LSTEAKSKDV-KIEVLQNELDVLQFSEQSTLIRLSQLOKNESEVL-----EG 850
QY 622 AFRKNHLKVKGSSSTLSQITFNAFDQNDGDKRFKKVVRTHNNSHMLADGERFIIDPK 681
Db 851 AERVRHI-----SSKVEEL-----SQALSQKELEITTKMDQ 880
QY 682 LSTMTVDLEKVLKIEGLDYSLRQTQTQFREMTTAGKALLSKSNLRQGEKQKFLQEAQ 741
Db 881 LL-----LEK-----KRDVETLQQTIEEKDQQT-----EISFSMTEKMWQVINEEK 921
QY 742 PFLGRVLDKAIKAKEKALVTKATKNGHLLERSINKAVLAYNSAIIKA--NVKRLK 798
Db 922 FSLG-VEI-KTLKEQLNLISRAEAKKEQVEEDNEVSGLKQNYDEMSPAQISKELOH 979
QY 799 ELDLLITDLVEKGPLAQATWQGVLLKTPLPPEVYIGLVNYFDKSGKLIVYALDMSDT- 857
Db 980 EFDLLKKNENQKRKLQAAIINRKELLQVSRLEBELANLK--DESKK--EIPUSETE 1033
QY 858 IGBGQKDAYGNPILNVEDENEGYHTLAVATLADYEGLYTKDILNSSLDKIKAIROPLAK 917
Db 1034 RGEVEE-----DKENKEYSEKCVTSKQCEIETIKOTISEKEVELQHIRKOLEEK 1083
QY 918 YHRLGIFQAIRMAAEADRLLPKTPKGYLNEVPNRYKKQMEKNLKPVDYKTPINFALPN 977
Db 1084 L-----AAEEQFQALVK---QMNQTLQDKTNQIDLLQAEISENQAIIOKLITS 1128
QY 978 --EKVDGDRAKAGHNINAEITNSVAVTP 1003
Db 1129 NTDASDGDGVAL-----VKETVVISP 1149

RESULT 13
KTNI HUMAN
ID KTNI HUMAN STANDARD; PRT; 1357 AA.
AC Q86UP2; Q13999; Q14707; Q15387; Q86W57;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Kinectin (Kinesin receptor) (CG-1 antigen).
GN KTNI OR CGI OR KIAA0004.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lymphoid;
RX MEDLINE=95306853; PubMed=7787243;
RA Fuechterer A., Kruppa G., Kraemer B., Lemke H., Kroenke M.;
RT "Molecular cloning and characterization of human kinectin.";
RL Mol. Biol. Cell 6:161-170 (1995).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Peripheral blood lymphocytes;
RX MEDLINE=94314220; PubMed=8039706;
RA Print C.G., Leung E., Harrison J.E.B., Watson J.D., Krissansen G.W.;
RT "Cloning of a gene encoding a human leukocyte protein characterised by
RT extensive heptad repeats.";
RL Gene 144:221-228 (1994).
[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Wang H.-C., Chen W.-F., Su Y.-R.;
RT "Identification of a variant of Homo sapiens kinectin mRNA.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.";
RL DNA Res. 1:27-35 (1994).
[5]
RP SEQUENCE FROM N.A.
RX PubMed=12508121;
RA Heilig R., Eckenberg R., Petit J.-L., Fontknechten N., Da Silva C.,
RA Catolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,
RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,
RA Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C.,
RA Brule T., Jaillon O., Friedlander L., Samson G., Brothier P.,
RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,
RA Gouyvenoux M., James R., Madan A., Maltrey-Batrada B., Mangenot S.,
RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Traak B.,
RA Vacherie B., Bellemere C., Blierz-Silla S., Besnard-Gonnet M.,
RA Bartol-Mavel D., Boutard M., Brierz-Silla S., Combette S.,
RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D.,
RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,
RA Dumont C., Guerin T., Halfray S., Hammadi R., Muanga J., Pellouin V.,
RA Robert D., Wunderle E., Gauguier G., Roy A., Sainte-Marthe L.,
RA Verdier J., Verdier-Diecala C., Hillier L.W., Fulton L., McPherson J.,
RA Matsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W.,
RA Quetier F., Waterston R., Hood L., Weissbach J.;
RT "The DNA sequence and analysis of human chromosome 14.";
RL Nature 421:601-607 (2003).
[6]
RP SEQUENCE OF 1-870 FROM N.A. (ISOFORMS 1/2).
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Caavaant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN 171
RP SEQUENCE OF 191-195; 395-405; 545-554; 747-761 AND 811-823.
RX MEDLINE=21969647; PubMed=11973345;
RA Tran H., Pankov R., Tran S.D., Hampton B., Burgess W.H., Yamada K.M.;
RT "Integrin clustering induces kinectin accumulation.";
RL J. Cell Sci. 115:2031-2040(2002).
RN [8]
RP CHROMOSOMAL LOCATION.
RX MEDLINE=96163023; PubMed=8575822;
RA Print C.G., Morris C.M., Spur N.K., Rooke L., Krissansen G.W.;
RT "The CG-1 gene, a member of the kinectin and ES/130 family, maps to
human chromosome band 14q22.";
RL Immunogenetics 43:227-229(1996).
CC -1- FUNCTION: Receptor for kinesin thus involved in kinesin-driven
CC vesicle motility. Accumulates in integrin-based adhesion complexes
CC (IAC) upon integrin aggregation by fibronectin.
CC -1- SUBUNIT: Parallel homodimers formed between the membrane-bound and
CC the cytosolic form, and also between 2 cytosolic forms (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Vesicle membrane
CC protein anchored to the endoplasmic reticulum.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q86UP2-1; Sequence=Displayed;
CC
CC Name=2;
CC IsoId=Q86UP2-2; Sequence=VSP_007981, VSP_007982;
CC -1- TISSUE SPECIFICITY: High levels in peripheral blood lymphocytes,
CC testis and ovary, lower levels in spleen, thymus, prostate, small
CC intestine and colon.
CC -1- SIMILARITY: Belongs to the kinectin family.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22551; CA80271.1; -;
DR EMBL; U25616; AAB65853.1; -;
DR EMBL; AY264265; AAP20418.1; -;
DR EMBL; D13629; BAA02794.1; -;
DR EMBL; AL138499; -; NOT ANNOTATED CDS.
DR EMBL; BC050555; AAH50555.1; ALT_TERM.
DR PIR; S32763; S32763.
DR PIR; I53799; I53799.
DR Genew; HGNC:6467; KTN1.
DR MIM; 600381; -;
DR GO; GO:0005789; C:endoplasmic reticulum membrane; TAS.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0006899; P:nonspecific vesicle transport; TAS.
DR GO; GO:0007018; P:microtubule-based movement; ISS.
DR InterPro; IPR02017; Spectrin.
KW Endoplasmic reticulum; Signal-anchor; Transmembrane; Coiled coil;
KW Alternative splicing; Polymorphism.
FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 7 29 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT DOMAIN 30 1357 ENDOPLASMIC RETICULUM (POTENTIAL).
FT DOMAIN 330 1356 COILED COIL (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 904 904 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1088 1088 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1263 1263 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1329 1329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 1031 1059 Missing (in isoform 2).
FT VARSPPLIC 1232 1259 /FTId=VSP_007981.
FT VARSPPLIC 282 282 Missing (in isoform 2).
FT VARIANT 282 282 /FTId=VSP_007982.
FT CONFLICT 15 15 V -> M (in dbSNP:2274073).
FT CONFLICT 210 210 S -> P (IN REF. 4).
FT CONFLICT 373 373 I -> M (IN REF. 1).
FT CONFLICT 939 939 E -> G (IN REF. 1).
SQ SEQUENCE 1357 AA; 156274 MW; 971FCDF8AA8FC88E CRC64;

Query Match 3.7%; Score 199; DB 1; Length 1357;
Best Local Similarity 19.2%; Pred. No. 0.062;
Matches 218; Conservative 210; Mismatches 404; Indels 302; Gaps 59;

QY 2 TKKHILKTLALATTVS-----VITYSQEVYGLEREESVQKQTSASEDDW-----FEE 50
DB 351 TKDRCKQLTQEMTEKERSNVVITRMKDRIGTLEKEHNVQNKIHVSYOETQOMKFOQ 410
QY 51 DNERK-----TNVSKENSTVDETVSDLFSDGNSNNSSTESVSDPKQVPAKPEVTQEA 106
DB 411 VREQWEAIAHLKQENGILRDVSN-----TTNQLSKQSAEINKLRQ---DYARLVNEL 462
QY 107 SNSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHTVL 166
DB 463 TEKTKLQOEVEVQKNAEQAAATQLKVLQEA----- 494
QY 167 PSHAADQTLTQVASFATPDKKTAIAEYTSRLGNGKPSRLDIDQKEIIDEGEIFNAYQ 226
DB 495 -----RRWEVQSYI---RKRTAEHE-----AAQODLSKFAKENEVQSLH- 533
QY 227 LTKLTIPNGYKSIGODAFVDNKNIAE-----VNLPSELETISDYAPAHMSLKQ 274
DB 534 -SKLT-----DTLVSKQQLQGRIMQLMSESKRVNKESLQ-----MQVQD 573
QY 275 VKLPDNLKYLIGELAFEDNQIGKL---YLPHILKLAERAFKSNRI-QTVEFLGSKLKVI 330
DB 574 I-LEQNEALKAQIQQFHSQIAAQTSASVLAELHKVI--AEKQKQIKQTEDSLASERDRL 630
QY 331 --GEASFQDNNLRNWMLPDGLEKIESEAFGTGNPGEHYNQVVLRTGTQGNPHQL--ATE 386
DB 631 TSKEEELKDIQNNMFLLKAFVQKQLALA-----NEQAAA-----AHELEKMQQ 673
QY 387 NTYVNPDKSLMRATPDMDYTKWLEEDFTYQ-KNSVTGF-----SNKGLOKVRANKLEIP 440
DB 674 SVYVKDDK-----IRLLEEQLQHEISNKMEEFKILNDQNKALKSEVQKLQTLVS 722
QY 441 KQHGNTITTEIGDNAFRNVDFOSKTLRYDLE--IKLPSTIRKIGAFQSNNLKSFEA 498
DB 723 EQPNKDVEEQM-EKCIQEKDEKLKVEEL-LETGLIQVATKEEELNARIENSSL----- 775
QY 499 SEDLEEIKEGAFMNNRIGTLDLKOKLIKI---GDAAFHINHIAVLPEVQEIGRSAFR 555
DB 776 TKEVQDLK-AKQNDQVSFASLVEELKVIHEKDGI-----KVEELLEA--- 819
QY 556 QNGALHLMFIGNKVKYTI---GEMAFSLNKLSEVNLSEQOKL-TIEVQAFSDNALSEV 610
DB 820 -----ELTKVANKERTVQDLKQEIFALKKEIGNVQLEKAQQLSTISKVQELQ----- 866
QY 611 LPPNLQTIREEAFKRNHLKEV-----KGSSTLSQITFNAFDQNDGDKRGKVVYRTHMN 665
DB 867 ---NLKKEE--QNMNTKAVLEKEKDLATGKWLQDLQEENESLKAHVQEV--AQHN 918
QY 666 SHMLADGERFIIDPKLSTWMDLEKVLKIEGL-----DYSTRQTTQOTQRFEMTTAG 719
DB 919 LKEASSASQF-----ELEIVLKEKENELKRLAAMLKERESDLSKSTQLLQ---DVQDEN 970
QY 720 KALLSKSNLRQGEKQFLQEAQFGLGRVLDKAIKAKEKAL-----VTKKAT 766

```

Db      971 K--LFKSQIEQLKQONV--QQASSFPPEHLLKVTSEREKEISGLMNELDSLKDAVEHQRK 1027
QY      767 KNGHLLERSINKAVLAYNNSAIKANKVRLKEKELDLTLVEGKGPLAQMVGVTLLK 826
Db      1028 KNDLREKNWEAMEALASTEKLQDKVNKTSKERQQOVEAVE-----LEAKEVLKKLPPK 1082
QY      827 TPLP-----LPEYIGLVNVPDYSGLIYALDMSDTIGEGQDAYGNPILNVEDNEGYHT 882
Db      1083 VSVPSNLSYGEWLHG-----PEKKAKEC---MAGTSGSEEVKYLEHKLKEADE---MHT 1130
QY      883 LA-----VATLADYEGLYIKDILNSSLDKIKAIRQIPAKYHRL--GIFQAIRMAAE 933
Db      1131 LLOECEKYSVLAETEGILQK--LQRSVEQEEKNKVKYVDESHKTIKQMOSSFTSSEGE 1188
QY      934 ADRLLEKTPKGYLNEVPNPKKQ--MEKNLKPDV-----YKTPI-----FNKALPN 977
Db      1189 LERLRSEN-----KDIENTLRREHLEMELEKAEEMERSTYVTEVRELKDLTLTELQKLLDD 1243
QY      978 EKYDGDRAAKGHN--INAETNNSVAVTPIRSEQ-----QLHKSQSDVNLPGT 1022
Db      1244 SYSEAVRQNEELNLKAQOLNET--LTKLRTQNERQKVAQDLHKAQOSLELIQS 1295

```

RESULT 14

```

SPOF_SCHPO STANDARD; PRT; 1957 AA.
ID SPOF SCHPO
AC Q10411; Q9USE9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SPO15 OR SPAC1F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
RL to the spindle pole body and essential for its modification.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymompres B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT      "The genome sequence of Schizosaccharomyces pombe.";
RL      Nature 415:871-880(2002).
RN      [3]
RP      SEQUENCE OF 705-871 FROM N.A.
RC      STRAIN=968 h90;
RX      MEDLINE=20223868; PubMed=10759889;
RA      Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Harauchi T.,
RA      Hiraoaka Y.;
RT      "Large-scale screening of intracellular protein localization in living
RT      fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL      Genes Cells 5:169-190(2000).
CC      -!- FUNCTION: Has a role in the initiation of spore membrane
CC      formation.
CC      -!- SUBUNIT: Monomer.
CC      -!- SUBCELLULAR LOCATION: Spindle pole body.
CC      -!- SIMILARITY: Belongs to the MPC70 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z70690; CAA94624.1; -.
DR      EMBL; AB027811; BAA87115.1; -.
DR      PIR; T38077; T38077.
DR      GeneDB; Spombe; SPAC1F3.06c; -.
KW      Sporulation; Coiled coil.
FT      DOMAIN 199 785 COILED COIL (POTENTIAL).
FT      FT 804 1235 COILED COIL (POTENTIAL).
FT      FT 1320 1471 COILED COIL (POTENTIAL).
FT      FT 1481 1723 COILED COIL (POTENTIAL).
SQ      SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

```

Query Match 3.7%; Score 199; DB 1; Length 1957;
 Best Local Similarity 19.4%; Pred. No. 0.1;
 Matches 227; Conservative 204; Mismatches 449; Indels 292; Gaps 54;

```

QY      2 TKHKLTLALATTVSVTVYSGEVYGLERESVQEQTSASEDDWFEEDENRKTVNSKE 61
Db      575 SKNSLQTLCAFAQEKAKSVWQ-----LKENQNFSSLDTSFKKLNESHQLENN 624
QY      62 NSTVDETVDLFSQDGNSSSKTESVSDPKQVPKAKPEVTQEAASNSNDASKVEVPKQ 121
Db      625 HQITTKQLKD-----TSSKLQOL-----QLERANFEQKESTLSDENNDIRTKLIKL 670
QY      122 DTASKK-----ETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPASHADG 173
Db      671 ESNKSLIKQEDVDLSLEKNIQTLKEDLRKSEALRFSKLEAKNLREVIDNLKGHEITL 730
QY      174 TQLTVASFAFTPPDKTAIAEYTSRLGENGKP--SRLLDIDQKEIIDE-----IFNA 224
Db      731 AQNDLHS--SLSDAKNTNAIILSSLETKSSSEYVKRLTANVETLTQDSKAMKQSFSLVNS 788
QY      225 YQLTKLTIPNGYKSIQD-AFVDNKNIAEVLNPESETISDYAFAMMSLKQVKLPDNLK- 282
Db      789 YQ---SISNLVHELRDDHVNMQSQNTLLSESKLTDCE---NLTQQNMTLLDNVQK 840
QY      283 -----VIGELAFPDNQIGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGE 332
Db      841 LMKHVNGESKVSSEL-----KEVNGKSLDLKLNRLSSLNVAISDNDQILTQLAELSK---- 892
QY      333 ASFQDNNLRNVMPLDGLKIESE-----AFTGNPGDEHHYNNQVVLRTRT 376
Db      893 -NYDSLQESAQINGSLKSLAEKQLHTENELHRLDKLTGKLKIEESKSSDLGKLT 951
QY      377 GQNPVH--QLATENTYVNDPKSLWRATPDMYT-----KMLEEDFTYQKNSVTFGSKGLQK 430
Db      952 ARQHEISNLKEEN--MSQQAITSVSKSLDETLSKSKLEADIHLKKNKVSVEVERNAL 1009
QY      431 VRANKNLEIPKQHNIGITI---TEIGDNAFRNVDFQSK-TLRKYDLEETIKLPSTIRKIGA 485

```

```

Db      1010 LASNERIMDDKNNNGENIASLQTEIEKKRAENDLQSKLSVSVSEYENLLISS-QTNKS 1068
QY      486 FAFQSNNLKSE-----ASEDLERI-----KEGAFMNRIGTL----- 518
Db      1069 LEDKTNQKYEKKNVOKLDEKQQRNVELBELTSKYKGLGENAQIKDELALRKSKQ 1128
QY      519 -----DLKDKLIKIDAAFINHIVAVLPESVQERISAFRONGALHLMFGNKV 569
Db      1129 HDLCANFVDLKEK-----SDALEQ-----LTNEKMLIVSLQSN-----SNNE 1168
QY      570 KITGENAFLSNKLESVNLSEQKQKLTIEVQAFSDNALSEVYLPENLQTIREEAFKRNHLK 629
Db      1169 ALVEERSDLANRLSDM-----KKSLS-----SDNVIS--VIRSDLVRVNDEL---DTLK 1213
QY      630 EVKGS-----STLSQTFNAPDQNG-DKRFQKVV-----VRTHNN-----SHMLADGER 674
Db      1214 KDKDSLSTQYSEVCCQDRDDLDSLKGCEBSFNKYAVSLRELCTKSEIDVPVSEILD--N 1271
QY      675 FIIDPDKLSS---TMVLEKVLKIEGLDYSTR---QTTQTFREMTAGKALISK 725
Db      1272 FVFNAGNFSLSRLTLVLSLENYLDAFNQVNFKMELDNRLTTDAEFTKVADLEKL--- 1328
QY      726 SNLRQGEKQFLQEAQFLGRVDDKAIAKAEKALVTYKA--TKNGHLE--RSINKAVL 781
Db      1329 ---QHEHDDWL-----IQRGDEKALKDSEKNFLRKEAEMTENIHSLEGEKETKEI 1378
QY      782 AYNSNAIK-----KANVRLKEKELDLTLTDLVEGKPLAQATMVQGVYLLKTPPLP 832
Db      1379 AELSSRLDNQLATNKLKNQDLHNLQERLKEVDLKEK----- 1416
QY      833 EYYTGLNRYFDKSGKLIYALDMSPTIGEQKDAYGNPILNVEDNEGYHTLAVATLADYE 892
Db      1417 -----ESLISLES--LSNQOKES-----SLDAKNELEHMLDDTSRKNSS 1457
QY      893 GLYIKDILNSLDK-----IKAIRQI-PLAKYH--RLGTFQAIRNAABADRLPKTPKG 944
Db      1458 LMEKIESINSLLDKSFLASAVEKLGALQKLHSESLIMENTKSQLOEAKETI-QVDES 1516
QY      945 YLNEVP-----NYRKKOMEKN--LKPVDYKTPLENKALPNEKVDGDRAKGHINAE 994
Db      1517 TIQELDHEITASKNNYEGKLNKDSITRDLSENIEQLNNLLAEK---SAVK--RLSTE 1570
QY      995 TNNSVAVTPIR-SEQQLHKSQSDVNLPTSSK 1025
Db      1571 KESEILOFNSRLADLEYHKSQVESSELSGRSKLK 1602

```

RESULT 15
USOL_YEAST STANDARD; PRT; 1790 AA.

```

ID      USOL_YEAST STANDARD; PRT; 1790 AA.
AC      P25386;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Intracellular protein transport protein USOL.
GN      USOL OR INT1 OR YDI058W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=X2180-1A;
RX      MEDLINE=91185402; PubMed=2010462;
RA      Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA      Yamaaki M.;
RT      "A cytoskeleton-related gene, usol, is required for intracellular
RT      protein transport in Saccharomyces cerevisiae.";
RL      J. Cell Biol. 113:245-260 (1991).
RN      [2]
RP      SEQUENCE OF 782-1790 FROM N.A.
RA      Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,

```

```

RA      Kendrick K.B.;
RL      Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 1-8 FROM N.A.
RA      Bai Y., Symington L.S.;
RL      Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: Required for protein transport from the ER to the Golgi
CC      complex.
CC      -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC      MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC      ER AND THE GOLGI COMPLEX.
CC      -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed
CC      of an heptapeptide repeat pattern characteristic of alpha-helical
CC      coiled coils. May form filamentous structures in the cell.
CC      -!- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X54378; CA38253.1; -
DR      EMBL; L03188; AAB00143.1; -
DR      EMBL; U53668; AAB66659.1; -
DR      GenOnline; 140300; -.
DR      SGD; S0002216; USOL.
DR      InterPro; IPR008938; ARM.
DR      InterPro; IPR002017; Spectrin.
DR      InterPro; IPR006955; USOL_P115_C.
DR      InterPro; IPR006953; USOL_P115_HEAD.
DR      Pfam; PF04871; USOL_P115_C_1.
DR      Pfam; PF04869; USOL_P115_HEAD_1.
DR      K01 Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
KW      DOMAIN 1 724 GLOBULAR HEAD.
FT      DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT      DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT      DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT      DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT      CONFLICT 847 847 G->E (IN REF. 2).
FT      CONFLICT 924 924 E->K (IN REF. 2).
FT      CONFLICT 1253 1253 V->I (IN REF. 2).
FT      CONFLICT 1319 1319 I->V (IN REF. 2).
FT      CONFLICT 1461 1461 N->S (IN REF. 2).
FT      CONFLICT 1581 1581 G->S (IN REF. 2).
FT      CONFLICT 1600 1600 I->V (IN REF. 2).
FT      CONFLICT 1661 1661 R->S (IN REF. 2).
FT      CONFLICT 1772 1772 D->DEEDDEE (IN REF. 2).
SQ      SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

```

Query Match 3.7%; Score 198.5; DB 1; Length 1790;
Best Local Similarity 19.3%; Pred. No. 0.096;
Matches 232; Conservative 209; Mismatches 432; Indels 329; Gaps 57;

```

QY      16 VSVVTYSQEVYGLERE-ESVQEQTQASASDDWFEED-NERKTNVSKENSTVDETVSDLF 73
Db      724 ISKISF-EEVEKLGQRQCTKLKEGITSLOTETESTHENLTKLIALTNEHKLDEKYQIL- 781
QY      74 SDGNSNNSSKTESV--SDPKVPPKAPVETQ-----EASNSNDASKVE---VPKQD 122
Db      782 ---NSHSSLIKENSFILELKNVRDSLDEMTQLRDVLETKYKENTALLLEYKSTIHKE 838
QY      123 TASKK-----ETLETSTWEAKDFVTR-GDTLVGFS-----KSGINKL 158
Db      839 DSIKLEKLEETILSQKKAEDGINKGKDLFALSREMOAVENCKNLQKEKDSNVNHQ 898
QY      159 SQTSHLYLPASHADGTQLTQVASFATPDKTAIAEYT--SRUGENGKPSRLDID----- 211
Db      899 KETKSL-----KEDIAAKITEIKAINENLJEMKIQCNNLSKE 935
QY      212 ----QKEIIDGEIIFNAYQ--ILTKLT-----IPNGYKSIGQD-----AFVDKNIAEV- 253

```


Db 936 KEHISKELVEYKSRFQSHDNVLAKITTEKLSLANNYKMQAENESLIKAVEESKNESSIQ 995
QY 254 --NLPESETIS-----DYAFAMSLKQVKLPDNLKVIIGELAFPDNQIGK----- 297
Db 996 LSNLQNKIDMSQEKENFQIERGSIENKIEQK---KTISDLQOTKEEIIKSDSSKD 1050
QY 298 -----LYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGE-ASFQDNMLRNVMLPDGLEK 351
Db 1051 EYESQISLKEKLTATANDENVNKISELTKTRELEAFLAAYK--NLKN-ELETKLEJT 1107
QY 352 IESEAFNGPDEHYNNQVVLRTKGQNPQOLATENTYVNPDKSLMRATPDMDYTKWLEB 411
Db 1108 SEKALKEVKEHEHLKEEKI-----QLEKBAETKQQLNSLRA--NLESLEKEHE 1155
QY 412 DFTYQ-KNSVTGFSNKGLOKVRNKNLPIPKQNHGITTIEIGDNAFR---NVDPOSKTL 466
Db 1156 DLAQOLKKEEQIANK-----EFOYNEISQUNDEITSTQOENESIKKQNDLEGEVKAM 1210
QY 467 RKYDLEIKLPSTIRKIGAFQSNLKSFEASED---LEIKGAFMNNRIGTIDLKDK 523
Db 1211 KSTSEEQSNLKKS--EIDLNLQIKELKKNETNEASLESIK-----SVESE 1256
QY 524 LIKIGDAFHINHITAIVLBESVOEIGRSAPRONGALHMFIGNKV--TIGEMAFLSNK 581
Db 1257 TVKIKELQDECNFKEKEV--SELEDKLKASEDKNSKYLELOKESEKIKEELDAKTEELKIQ 1315
QY 582 LESV-NLSEQOKLTIEVOAFS-----DNALSEVVLPNLQITIREAFKRNHLKEVG 633
Db 1316 LEKITNLKAKEKESSELSRLKTSSEERKNAEQLEKNEIQIKQAFEKERKLINEG 1375
QY 634 SSTLSQITFNAFDQNDGKRFQKV-----VVRTHNNSHMLADGERFIIDPKLSSTMV 687
Db 1376 SSTITQ-----EYSEKINTLEDELIRLQNMENELKA-----KEIDNTRS 1413
QY 688 DLEKVLKIEGLDYSTLRQTTQTFR-----EMTTAGKALLS--KSNLRQGEKQ 734
Db 1414 ELEKV-----SLSNDELEEKONTIKSLQDEILSYKDKITRNDKILSTIERDNKRDLSEL 1468
QY 735 K-FLQEAQFLGRVDLKAIAKAKAL--VTKKATQNGHLLERS-----INKA 779
Db 1469 KEQLRPAQ-----ESKAVEEGCLKLEBESSKEKALEKSKEMMKLESTIESNET 1519
QY 780 VLAYNNSAIKKA-----NKRLEKE-LDILLDLVEGKPLAQATVQGVYLL 825
Db 1520 ELKSSMETIRKSDEKLEQSKSAEEDIKNLQHEKSDLSRINSEKDIIE-----L 1570
QY 826 KTPPLPEYIYIGLNVYPRKSGKLIYALDMSDTIGEGQDAYGNPILNVDEDENEGHTLAV 885
Db 1571 KSKLRIEA-----KSGSEL-----ETVKQELNNAQEKIRINAE--NTVLK 1609
QY 886 ATLADYE-----GLYIKDILNSLSDKIKAIRQIPLAKYHRLGIFQAIRNAAAE 933
Db 1610 SKLEDIERBLKDKQAEIKSQEKEKELLTSRLKELE-----QE 1646
QY 934 ADRLLPKTPKGYLNEVPNVRKKOMEKNLKPVDYKTPIFNKALBNEKVDGDRAAKGHNINA 993
Db 1647 IDSTQQAQKSEERRAEVRFQVEKS--QLEKAMLEETKY--NDLVNKEQAMK----- 1697
QY 994 ETNNSVAVTPIRSEQQLHKSGQDV-NLPQTSKNNFIYEILIGVSLCLFLVTAGKKGR 1052
Db 1698 RDEDTVKTTDSQROEIEKLAELDNLKAENSKLKEANEDRSEIDDIMLVTDLDEKNAX 1757
QY 1053 AR 1054
Db 1758 YR 1759

Search completed: April 27, 2004, 11:15:52
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 11:11:11 ; Search time 27 Seconds
(without alignments)
3758.595 Million cell updates/sec

Title: US-10-091-007A-24

Perfect score: 5352

Sequence: 1 MTKKHLKTLALTTVSVT.....VSLCLLFLVTAGKKGRARK 1055

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pirl:*
2: pirl2:*
3: pirl3:*
4: pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	331.5	6.2	1081	2 T31094	surface antigen Bs
2	234	4.4	1127	2 T28317	ORF MSV156 hypothe
3	224	4.2	2748	2 S57976	nuclear migration
4	223.5	4.2	1252	2 B42771	reticulocyte-bind
5	216	4.0	2139	2 T18296	myosin heavy chain
6	215.5	4.0	1727	2 T50073	myosin-like coiled
7	212.5	4.0	621	2 A95250	choline binding pr
8	212.5	4.0	2401	2 T28676	rhodopy protein -
9	212.5	4.0	2819	2 A90551	conserved hypothet
10	209	3.9	1208	2 T39068	coiled coil protei
11	209	3.9	1875	2 S38173	myosin-like protei
12	209	3.9	6713	2 B89921	hypothetical prote
13	208.5	3.9	1227	2 C97033	uncharacterized pr
14	207.5	3.9	1119	2 B70126	surface-located me
15	207	3.9	2269	2 T28677	rhodopy protein -
16	203	3.8	690	2 F98114	choline-binding pr
17	202.5	3.8	2829	2 A42771	reticulocyte-bind
18	202	3.8	1940	2 A59287	myosin heavy chain
19	201.5	3.8	1365	2 T30822	lmpl protein - Myc
20	200	3.7	872	2 S55940	telomerase compone
21	199.5	3.7	1072	2 A86827	hypothetical prote
22	199.5	3.7	3225	2 I52300	hypothetical prote
23	199.5	3.7	3259	1 A56539	giantin - human
24	199	3.7	1272	2 C90593	hypothetical prote
25	199	3.7	1385	2 D89824	hypothetical prote
26	199	3.7	1957	2 T38077	hypothetical coile
27	198.5	3.7	1173	2 T43527	sp8 protein - firs
28	197.5	3.7	2166	2 G70163	hypothetical prote
29	197	3.7	1790	2 S67593	transport protein

30	195.5	3.7	1300	2 I53799	CG1 protein - huma
31	195.5	3.7	1992	1 S02771	myosin heavy chain
32	195.5	3.7	2712	2 T05113	hypothetical prote
33	195	3.6	1356	2 S32763	kinectin 1 - human
34	194.5	3.6	1864	2 F86378	protein F21J9.12 [
35	192.5	3.6	1358	2 A29360	SIR4 protein - yea
36	192.5	3.6	1839	1 OYBYK	adenyrate cyclase
37	192.5	3.6	1959	2 AG1085	hypothetical prote
38	192	3.6	1026	2 C97783	hypothetical prote
39	190	3.6	1132	2 H82887	cell surface anti
40	190	3.6	1558	2 B71603	hypothetical prote
41	190	3.6	4919	2 T31105	RESA-H3 antigen PF
42	188.5	3.5	1837	2 T41023	hypothetical prote
43	188.5	3.5	4688	2 F82885	probable nuclear p
44	188	3.5	719	2 A81358	hypothetical prote
45	188	3.5	1302	1 JC6009	surface-located me

ALIGNMENTS

RESULT 1

T31094
surface antigen Bspa - Bacteroides forsythus
C/Species: Bacteroides forsythus
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T31094
R/Sharma, A.; Sojar, H.T.; Glurich, I.; Honma, K.; Kuramitsu, H.K.; Genco, R.J.
Infect. Immun. 66, 5703-5710, 1998
A/Title: Cloning, expression, and sequencing of a cell surface antigen containing a le
A/Reference number: Z20977; MUID:99043895; PMID:9826345
A/Accession: T31094
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1081 <SHA>
A/Cross-references: EMBL:AF054892; NID:g3005672; PID:g3005673; PIDN:AAC82625.1
C/Genetics:
A/Genes: bspa

Query Match	Query Match	Score	DB 2	Length	1081;
Best Local Similarity	23.2%	Pred. No. 8.8e-08;			
Matches	173;	Conservative	99;	Mismatches	239;
Indels	235;	Gaps	32;		
QY	63	STVDETVSDLFSDGNSN--NSSSKTESVSDPKQVPKAKPEVTQEAAGNSNDASKVEVP	119		
DB	18	TTLGATAQNSGTTGPLNWSYDSGTTLATG-----TGAMPDF-----NNAS--EIP	62		
QY	120	KQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHTVPS-----HAAD	172		
DB	63	WHSLSQSKIQT-----VTIGDGVTSVGNNAFSDCALTS-VTLPNSLTAIGDHAFF	110		
QY	173	G-----TQLTQVASFAP-----TPDKKTAIAEYTSRLGENGKPSRLDIDQ	212		
DB	111	GCSGLTSTIPNSVTTIGEMAFKCGSLKSTLTPNSLTAIGQ-SALSGCTGLTSTIPNS	169		
QY	213	KEIIDEGEIFNAYQITKLITIPNGYKSGIDAFVDNKNIAEVLPESETISDYAFAMS-	271		
DB	170	VTTIGEMAFKCGSLTSTIPNSLTAIGESAFYCGALTSITLPPDALTTIGESAFKCGSG	229		
QY	272	LKQVLPDNLKVIGELAFPDNIGGKLYLPHLLIKLAERAFKS-NRIQVYFGLSKLKI	330		
DB	230	LKSITFPNSLTIGESAFYDCALTSITLPPDALTTIGRSAPYCGSGSLKSTIPNS-LTTI	288		
QY	331	GEASFQD-----NNLRNMLPDGLEKISEAFTGNPGDEHY	366		
DB	289	GESAFYNGSLTSTIPNSVTTIGRSAPYCGSGSLKSTLPPDGLTTIEBAF-----Y	340		
QY	367	NNQVLRTRTGQNPQLAT--ENTYVNPD--KSLMRATPD-MDYTKWLEEDFTYQKNSVT	421		
DB	341	NCGLVTSITI--PNSVATIGESAFYCGSGLSKI--TLPPDGLTTIEW-----G	383		
QY	422	GFSNKGLOKVRNKNLEIPKOHNGITITTEIGDNAFRN-----VDFQSKTLR	467		

Db 384 AFVNCGL-----TSTIPN-----SVSTIGESAFYGCALKDVTAMDPIDIGRDYR 433
Qy 468 KYDLBEIKL-----PSTI 480
Db 434 ELTSGIRLHVPAKKTVEAKDVKEFNIVEDDFGGLQWYDAATKTLTITNPTPTP 493
Qy 481 RKIGAFAFQSNLKSFEASEDLBEIKEGAFMNNRIGTLDLKLKIGDAFH-INHIYA 539
Db 494 KPMNPATPNDQJ-----WGAFF-ÖKEIÖKITIGDVTSGVDFAFSGCDALKS 539
Qy 540 IYLPESVOEIGRSAFRONGALHLMFIGNKVKITIGEMAFLSNKLSEVNLSEÖKÖLKTIEVÖ 599
Db 540 ITLPKSVTTIGQSAFSGCWDLSLTPDGVNTIGEKAFY-DCLELTSITIPKSVTALGQE 598
Qy 600 AF-----SDNALSEVVLPPNLTQIREEAFKRNHLKEVKGSS 635
Db 599 TFHYCVSLSLTLPLDALTAIGKKAFFYSCNALTSVTFPKSITIGENAF-----DGCT 650
Qy 636 TLSQITFNAFDQNDGDKRFGKKVVR 661
Db 651 ALKDLTV-AWKDNASIPDIGSKDVEK 675

RESULT 2

T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C/Species: Melanoplus sanguinipes entomopoxvirus
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C/Accession: T28317
R./Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A/Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A/Reference number: Z20484; MUID:99102612; PMID:9847359
A/Accession: T28317
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1127 <AFO>
A/Cross-references: EMBL:AF063866; MID:g4049647; PIDN:AAC97677.1; PID:g4049717
C/Genetics:
A/Note: MSV156

Query Match 4.4%; Score 234; DB 2; Length 1127;
Best Local Similarity 18.2%; Pred. No. 0.0032;
Matches 21; Conservative 189; Mismatches 346; Indels 416; Gaps 56;

Qy 19 VTYSQEE-VYGLERESVQEQÖTQASSEDWFEEDNERKTNS--KENSTV---DETVS 70
Db 108 ITYKKKLYIDLDYBEKKDKELVINIEÖKNAVVDKINDIKNNVNNIHSNMTIITGKETLI 167
Qy 71 DLFSGNSNNSKTESVSDPKÖVPKAKPEVTOEASNSNDASKVEVPQÖTASKEETL 130
Db 168 DIL-----NKLKLVSSDEKÖLIE--ÖIYKNNKEIEPKNIDNVÖKEINKÖDEL 216
Qy 131 ETSTWEA-KDFVTRGDITLVGFSKSGINKLSÖTSHLVLP SHAADGTÖLTQVAFATPDCK 189
Db 217 NKLLDESKKEFIKKÖEEL-----NK-----TIDCK 241
Qy 190 TATAEYSTRLGENGKPSRLDIDÖKE-IIDE-----GEIFNAYÖLTKLTIPNGY 236
Db 242 Ö--EELIKKL--NDKEINENIDEKÖKLDÖINSKINTLNENIKGVMLYETIKKISNLÖ 297
Qy 237 KSI-GÖDAFVD-----KNNAIEV-----NLPEBLE-TISDYAFAM 270
Db 298 NEILNKDSTIKSLDEKÖKLDELDKNINNITSLYKNSNTKITNIÖQLLESSLTDENNANI 357
Qy 271 SLKÖVKLPDLNLKVIIGELAFEDNÖIGKLYLPRHLIKLAERAFKSNRIÖTVEFLSKUKVI 330
Db 358 NINELK-----SKIKLFDNDIQ-----KLNNDIÖÖNN-KITDFENNSTRIF 398
Qy 331 GEA-----SFÖDNNLENNVMLPDGLEKIESEA-FTGNPGDEHYNNÖVVLRTTGÖNP 380
Db 399 KEKLDTEYKKIDIDIKNNNÖK--LEESYKKIDEÖTEYKKNKINKENYNDITIELK---NNNL 453

Qy 381 HÖLATENTYVNPDKSLMRATPDMDYTKWLEBEDFTYÖKNSVTGFSNKGLOKV--RRKNL-- 437
Db 454 ÖKLEENKID-----EÖTEYKKNKINKENYNDITIELKNNNLÖKLEENKNIND 501
Qy 438 EIPKÖHNGITITEIGNAFRNVDFÖSKT-----LRKYDLBEIKLPSTIRKIGA 485
Db 502 KLTILKNDIESNTLEFNKLNISDFKÖKSREIAKLNTYEÖLRKÖLLENINKTNELMKL-- 559
Qy 486 FAFÖSNLKSFEASEDLBEIKEGAFMNNRIGTLD-----LÖDKLIKIGDAFHIN--HIY 538
Db 560 ---SDNKLSSLEÖLYDSK-----NILDGIDKIYNSLKEKNDKIDEYFSNIEKEDIY 608
Qy 539 AIVLPESVOEIGRSAFRONGALHLMFIGNKVKITIGEMAFLSNKLSEVNLSEÖKÖLKTIEV 598
Db 609 NVI-----ENKFIÖNLDSIINKI--INNDÖFKEYINSKI 640
Qy 599 ÖAFSDNALSEVVLPPNLTQIREEAFKRNHLKEVKGSSTLISQITFNAFDQNDGDKRFGKV 658
Db 641 DS-----KSNEL-----STWFDDIFNAKNÖI-----ASI 664
Qy 659 VVRTHNSHMLADGERFIIDBDKLSSTMVD-----LEKVLKIEGLD-----YSTLRÖ- 706
Db 665 TNNIENISNKIKDLNEFIISNEDSSKEILLDEIRKYYKÖFDKIKDAMNTEVKSFEENTLÖKD 724
Qy 707 --TÖQÖREMTTAGKALLSKSNLRÖGEKÖKFLÖEAOÖFELGRVLDKALAKAEKALVTRK 764
Db 725 IDSIKSNINELTNAYDIINTKAN-----DLD-----DK 752
Qy 765 AITNGHLLERSINKAVLAYNNSA-----IKKANVKRLEKEJLDTLVEGKPLAÖATMV 819
Db 753 LNNYSEFEKN-----LYNNASDLLDTÖKNNDEKVKÖ----- 784
Qy 820 ÖGVYLLKTPLPPEYIIGLVYFDKSGKLIYALDMSDTIGEGÖKDAYGNPILNVDEDNEG 879
Db 785 -----LNEYIEKNKN--ÖSIEINDIYNNFIKELIK--FNNTEITNKS 821
Qy 880 YHTLAVATLADYEGLYIKDILNSSLDKI-KAIRÖIPLAKYHRLGIFÖAIRNAAAEADRL 938
Db 822 LNELT-----NDDINDKIFKLYKELNKI-----STNNLL 851
Qy 939 PKTPKGYLINEVPNRYKK-----ÖMEKNLKPVDYK-----TPIFN 972
Db 852 ----KIYKNEIDNVNNEKLSIVLENLÖFINSFLSIEFNÖGSITSHINFLNTLAGINDVLN 907
Qy 973 KALPNEKYDGDRAAKGH-NINAEFNNSVAVTPIRSEÖQLHKSÖSDV-NLPÖTSSKNNFY 1030
Db 908 KL--NLKIMADTTRRGDTNIRDEIKNÖISSENIKSÖKFNEKNEKÖULKLISFNDKLNKNY 965
Qy 1031 EILGYVSL-----CLLFLVTA 1046
Db 966 ISAGYTEYVNNIEHECLKLYIA 987

RESULT 3

S57976
nuclear migration protein NUM1 - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YD8358.06; protein YDR150w
C/Species: Saccharomyces cerevisiae
C/Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C/Accession: S57976; S19052; S17018
R./Murphy, L.; Richards, C.; Harris, D.
submitted to the EMBL Data Library, July 1995
A/Reference number: S57971
A/Accession: S57976
A/Molecule type: DNA
A/Residues: 1-2748 <MUR>
A/Cross-references: EMBL:Z50046; MID:g899393; PID:g899399; MIPS:YDR150w
A/Experimental source: strain AB972
R./Kormanec, J.; Schaaff-Gerstenschlaeger, I.; Zimmermann, F.K.; Perecko, D.; Kuentzel,
Mol. Gen. Genet. 230, 277-287, 1991
A/Title: Nuclear migration in Saccharomyces cerevisiae is controlled by the highly re
A/Reference number: S19052; MUID:92079907; PMID:1745235
A/Accession: S19052


```

Db 473 VKDQSNVYVNLNQITTEENLIVTEKNLNGIDSTITNIEGALKESKG-NYEI-----GF 526
QY 447 --TITEIGDNAFRNVDFQSKTLRKVDLEIKLPSTIRKIGAFAPQSN--LKSFEASEDL 502
Db 527 LEKLEIEIGKRLKVDITKKSINS-----TWGNFSSLFNNFDLNOYDFNKNI 573
QY 503 E-EIKEGAFMNNKIGITLDLKDKLKIGDAFHINHITVAIVLESVQIEGRSAFRONGAL 560
Db 574 NDYENKNGEHIYNEFEGLN-----KISENLRNA----- 601
QY 561 HLMFIGNKVKTIGEMAFLSNKLSEVNLSEQKQKTIEVQAFSDNALSEVLPNLTQIRE 620
Db 602 -----SENTSDYNSAKTIRLEAQKEV-----NLNKEE 630
QY 621 EAFKRNHLKEVKSSTLSQITENAFDQNDGDKRFGKK--VVVETHNNSHMLADGERFI 677
Db 631 EANK--YLRDVKKVESF-RFIFNMKESLDKINEMIKKEQLTVNEGHGNVKQJVENIKELV 687
QY 678 DPKLSSFTWVLEKVLKIEGLDYSTLRQTQTQOFREMTTAK-----ALLS 724
Db 688 DENNLSDILKQATGKNBEIQKTHSTLKNKAKTILGHVDTSAKYVGIKITPELATTELLG 747
QY 725 KSNLRQGEKQFLQEAQFLL-----GRVLDLKAIAKAEKAL-----VTKKAT 766
Db 748 DAKLKTAEKELKESKNNVLETEMNSKNTNEDLVHKNIQDAYKVALEILLASDEIDTKQK 807
QY 767 KNGHLLEERSIN--KAVL--AYNNS-----AIKKANVKRLEKELDLLTDLVEGK 810
Db 808 DSKLIEMGNQIYLKVLINQYKKNKISSIKSEKAVSVKIGNVSKKHSLSKIT----- 861
QY 811 GPLAQATMVQGVYLLKTPLPPEYVIGLVNYPDKSGKLIYALDMSDTIGEGQKDAYGNPI 870
Db 862 -----CSDKSYDNITIALEKQTEL-QNLNRSFTQEK 890
QY 871 LNVDEDN-----EGYHTLAVATLADYEGLYIK----- 897
Db 891 TINTNSDKLEKIKTDESLKNALKTLEGEVNAKASSDNHEHVQSKSEVPNPALEIEKE 950
QY 898 ----DILNSSLDK-IKAIQIPLAKYHRLGIFQAIRNAA-----AEADRLLPKTPKGY 945
Db 951 ETDIDSLNTALDELKKGRTCEVSRY-----KLKIDVTKEISDTELTINTIEKNVKA 1004
QY 946 LNEVPNRYRKQWMEKNLKPVDYKTPIFN-KALPN-EKVDGDRAAKGHN--INAETNSVAV 1001
Db 1005 L----AYIKKNYEDYQDVLTLNEHFNTKQVSNEHPTNEDKSNKSEELTKAVTDSKTI 1060
QY 1002 TPIRSEQQLHKSQSDVNLPTSSKN-----NFIYEILGYVSL 1038
Db 1061 SKLKGVIIEVNEVENTEMNTIESSAKIEALYNELKNKKTSLNEITYQTSNEVKL 1112

```

RESULT 5

```

T18296
myosin heavy chain - Entamoeba histolytica
C/Species: Entamoeba histolytica
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C/Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A/Reference number: Z18865
A/Accession: T18296
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2139 <GUI>
A/Cross-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C/Genetics:
A/Gene: mhca
C/Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

```

```

Query Match 4.0%; Score 216; DB 2; Length 2139;
Best Local Similarity 20.2%; Pred. No. 0.056;
Matches 260; Conservative 188; Mismatches 494; Indels 342; Gaps 58;

```

```

QY 3 KKHLLTALALTTSVVTYSQEVYGLEREESVQEQTSASEDDWFEEDNERKTVNSKEN 62
Db 883 KREAENALASATAKTGELEAKIQDLEDKISELSKLSAALDK--QELNLKLENLEEDK 940
QY 63 STVDETVSDLFSDGNSN-----NSSKT-----ESVSDP 92
Db 941 BELKETIDNLKGLDKDSKLKGEDLEVEITELNSQINTLNATVDKDKTIAEMQESIDEKE 1000
QY 93 KOVPEKAPPEV--TQASNNSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLV- 148
Db 1001 DETTKLKGDIKLLEEKDDLEQDRADVSA TKDIAKLNKITTICEADAKDEIAKLEQELE 1060
QY 149 ---GPSKSGINKLSQTSHTLVPSHAADGTQLTQVASFAPTPDKTAIAE-----YTSR 198
Db 1061 DEENKKNKDLTNELQQTQ---LKGTEKSLAAQVAATKASDERDTLSQNLNENKLTYN 1117
QY 199 LGE-----NGKPSRLDIDQKEIID-----EGEIFNAY-----QLTKLTPNGY---- 236
Db 1118 LTKTKADLEKKISGLKQDYEDLEDQNKIEGDLRNAQRKIKELDEITTKGADVSOYLQKQ 1177
QY 237 -----KSIGQDAFVDNKNIAE-----VNLPSL-ETISDYAFAMSLKQV 275
Db 1178 KEEVESQIAKQOEKEKAIQNDVKNKEKTIKEKELFIQSLQEKUDETEVEKEDEAKKKEI 1237
QY 276 -----KLPNL-----KVLGELAFDQNLGGKLYLPRHLI 305
Db 1238 EKEMKALQEEKENVSSKNSSTEDKCKLIEDNLKDTQCKLDDMT-ADNE----- 1284
QY 306 KLAERAFKSNRIQTVEFLGSKLYIGEASF-----QDNLRNVMPL-DGLEKIESEAF 357
Db 1285 KLKAKA-KDLEAQNLNEVQDNHEKAVADAELLNKKKAQSDKELNSLKALBALTKAKSV 1343
QY 358 TGNPGDEHYNNQVLRTRTQGNPHQATENTYVNPDKSLMRAPEMD----- 404
Db 1344 SKNKDSE--NEKALSEEIDQANEKIK-----NIQADLRKATADLQEANKEKAVEAQR 1395
QY 405 -----YTKWLE-----EDFTY-----QKNSVTGFSNKGK--QKVRNKN 436
Db 1396 DKLAVADNKKQVTKTLEIKARDEENFYKVENYEKVLKRKEADLEBANENLDIEKKDRMKE 1455
QY 437 LEIPKQHNGITITEIGDNF--RNVDFQSKTLRKYDLEI-----KLPTIR 481
Db 1456 KQVKKLEGELEKETDKLNAIAIEKDSITFAKKQSDADLBEINKTVEEHDEVAKLNTQIT 1515
QY 482 KI-----GAFAPQSNNLKSFSEASEDLEIKE-----GAFMNNRIGITLDLKDLIKIGDAAF 532
Db 1516 KLTRDNGSAEBELNELRS-KADKDKKKTISELEQVNLSESPVPGNADENEIKIRDA-- 1572
QY 533 HINHITVAIVLESVQIEGRSAFRONGALHLMFIGNKVK-TIGEMAFLSN----- 580
Db 1573 QIADINKALEMKGVQNNQLQATNKELKAKDNLDLTSKIEITENEMKLENAKKRLQDKDE 1632
QY 581 -----KLESVNLSEQKQKTIEVQA--FSDNALSEV-----LPPNLQITIREAF 623
Db 1633 ADKAVSEQTIKRGLEEEVKLTTEIQALKFQINAPSSVAQEEKQRLIESDIALKEQLE 1692
QY 624 KRN-----HLKEVKSSTLSQITENAFDQNDGDKRFGKVVVVRTHNSHM----- 668
Db 1693 QERTTAANAFAERKTIQAEIDEVKFNLEDTNQR--KLVAKNSENDAEIDSLKEEKA 1749
QY 669 LADG-ERFIIDPKLSTWVLEKVLKIEGLDYSTLRQTQTQOFREMTTAKALL--SK 725
Db 1750 LEDELEKITTDDNNKLSSEIDSLDRKNAL--LDSKSDSVSMKEKQDELKVTKDLETEK 1807
QY 726 SN-----LR-QGEKQFLQEAQFLLGRVLDLKAIAKAEKALVTKKATNGHLLEERSINKAV 780
Db 1808 KNHAETMRLKGRLEKAAAEVQVRLBALQKNLPLAQOEKA---KATQYRAADGELKSLM 1863
QY 781 LAYNNSAIKKANVK-RLEKELDLLTDLVEGKBLAQ--ATMVQGVYLLKTPLPPEYVIG 837
Db 1864 NELD-----DYVDQLDKAQDDLDADKEDELATLDDQKYKTVKQSVFDSRIQEMQEQLD 1916

```


QY	176	LTQVAS--FAFTPDKTAIAEYTSRLGNGKPSRLDIDQKEIIDEGEIFNAVOLTKLTI	233
Db	98	VTEVASQAFSYPDETGRIVYPS-----SITIP	126
QY	234	NGYKSIGODAFVDMKNIAEV-NLPESLETTISDYAFAMSLKOVKLPDNLKVIGETLAFPDN	292
Db	127	SSIKKIQKGFHGSKAKTIIFDKGSQLEKIEDRAFDESELEIELPASLEYIGTSAFSFS	186
QY	293	QIGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGEASFOD--NNLRNMLPDGLEK	351
Db	187	Q-----KLKLTFSSS-----SKLELISHAEFANLSNLEKLTLPKSVKT	225
QY	352	IESEAFGNPGEHNNQVVLRTRTGONPHQLATENTYVNPDKSLMRATPDMDYTKWLEE	411
Db	226	L-----GSNLFRLTSLKHVDVEEG-----NE	247
QY	412	DFTYQKNSVTG--FSNKGLOKV-----RRKNLEIPKQHGKITITEIGDNAERNYDFQSK	464
Db	248	SFA-----SVDGVLFSKOKTQLIYYPSQKNDSEYKTPKETKELASYSFNKSY-----	295
QY	465	TLRKYDLEIKLPSTTRKIGAFAF-QSNNLKSFEASEDLLEIKEGAFMNNRIGTLDKDK	523
Db	296	-----LKKLELNEGLEKIGTFAFADAIKLEELSPNSLETIERLAF-----YGNLELKE-	344
QY	524	LKIGDAFHINHIVALVPESVQEIGRSAFRONGALHLMFIGNKVKITIGEMAFISNCKLE	583
Db	345	-----LILPDNVKONFGKHAMVNGLPKLKSLTIGNNINSLSPSF-FLSGVLD	387
QY	584	SVNLSEQOKLTIBVQAFSDNALSEVVLPPNLQITREEA--FKRNHLKEVKGK--STLS	638
Db	388	S-----LKEIHIK-----NKSTEPSVKKDTFAIPETIVKYVTSEHIKIDVLSNLSSTN	435
QY	639	QITFNAFD	646
Db	436	DIVEKVD	443

RESULT 8
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C;Species: Plasmodium yoelii
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C;Accession: T28676; A45521
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A;Title: Comparison of two members of a multigene family coding for high-molecular mass
A;Reference number: Z20507; MUID:97077455; PMID:8920022
A;Accession: T28676
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2401 <SIN>
A;Cross-references: EMBL:U36927; NID:G1041784; PID:G1041785; PIDD:RAB41263.1
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A;Reference number: A45521; MUID:91101660; PMID:2270106
A;Accession: A45521
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 2260-2401 <KEB>
A;Cross-references: GB:M34281

```

Query Match          4.0%; Score 212.5; DB 2; Length 2401;
Best Local Similarity 19.7%; Pred. No. 0.097;
Matches 238; Conservative 195; Mismatches 452; Indels 321; Gaps 57;

QY      4 KHLKTLALATTVSVVTYSQAEVYGLERESVVKQEQTQASAEDDWFEEDNERKTNVSKENS 63
      || : : : : : ||| : : ||
Db      15 KHFNELIKKIAESYSILFFNYTL--LENLESV--NITYKESLDYFF-----NS 57

QY      64 TVDETIVSDLFSDGNSNNSSSKTESVSDPKQVPK-AKPEVTQEAASNSNDASKVAVPKQD 122
      :: : |||| : : : : : : : : : :

```

Db	58	LGKLLINKVUSDGNIIEENDIDNFDLSKPKNNFKLLEGELNGVFNKMWYKNNKNLDOF	117
QY	123	TASKETLETSTWEAKDFVTRGDTLVGFSKSGI-----NKLSTQSHLVLP SHAADGTQ	175
Db	118	KDTMKKILLIIOYMNEFKGLNDAWTKLKNEGISQKFVINNOQKQKFDKSTYDEKKEGFE	177
QY	176	LTQVASFAPFPDCKTAIAEYTSRLGNGKPSRLDI-----DQKEIIDEGEIFNAYOQT	228
Db	178	SLLEAKNWEKKLEIITELKKK---NEETVOLDIKIRELIKQJLKDIIIEQKIVNDLKE	234
QY	229	KLITPNGYKSTIGODAFVDNKNIAEVNLPESETISDYAFAHMSLKQVKLPDNLKVIGELA	288
Db	235	L-----NKKIKEIT--EKIE---YIKKAVDLKEIEKDNV-YIDE LA	270
QY	289	FPDNÖIGKLYPRHLIKLAERAFKSNRIQTVFEFLGSKLVIGEASFQDNNLRNVM.LPDG	348
Db	271	-----KEPPYQITKYIEK--KNEIYNTIKSDFDKI-YVGDIHQLYNEMFSVQÖSN	318
QY	349	LEKIE--SEAFI--GNPGEDEHYNNQVVLRTGQNP--HQLATEN---TYVNPDKSLW--RA	399
Db	319	IEHIENTEILITKTKIDNVYNNIOQMETETVSKHLKNIETNNKLSSETIIDIICYIGEI	378
QY	400	TPMDYTKWLEDEFTYQKNSVTGFSNKGQKVRNNKNLEIPKQNHGITTITEIGDNAFRNV	459
Db	379	TNELNKTL--EDF---KNKEKGLSNKIDEBYAKENVQNLVYKSN-----ILEI-----	420
QY	460	DFQSKTLRKXYDLBEIKLPSTIRKIGAFQSNMLKSFEASEDLBEIEKG-----AF	510
Db	421	-----KHYNDQINI-----DNIKEKAQÖNYDQFHEHMKITIPNEMKY	459
QY	511	MNNRIGTLDKDKLIKIGDAAFHINHIVLVPESVQÖIGRSAFRQNALHMF--IGNK	568
Db	460	QKPSIEIKIMKDEFLSKVNKYNDFDKVY---KEKVE-----SEHNKFTELTNK	504
QY	569	VKT---IGEMAFLSNKLSESVNLSEQÖKLTIEVQAFSDNALSEV-----VL	611
Db	505	IKTEVSDEIKKYENKFNDSKSLINETKKSIEEYÖNINTLKKVDYIKVCLNTNELITN	564
QY	612	PENLQITIREBAFKRNHLKEVKGSSTLQITENAFDQNDGDKR-----FGKVVVRTHN	664
Db	565	CHNKQÖTTLKDKLNÖN--IKTIKETNSIDKIYTDKFENILTDKKTLEFTKFTGLSLNNHESN	623
QY	665	NSHMLA-----DGERFIIDPDKLSSTMVBLEKYLKIEGLDYS	702
Db	624	NKELLTYFYDLKANLGNKNENMLYKQFNEKEKAVEDIKKN--VDINKIVSNIETIYT	680
QY	703	TURQÖTÖ-----TQFREMTTACKALLS--KSNLRQGEKQKFLQEAQFLLG-	745
Db	681	STYNINEDTENEIGKSTIELNTKVLKVKKAVTNLNEIKEKLKDYDFQDPGKEKNIKYPD	740
QY	746	----RVDLKAIAKAEKALVT---KKATKNGHL--LERSINKAVYLAYNNSAIKKANVR	795
Db	741	ENKIKNDIDITLNÖKIDXSIEITLTIKKNSEN-HIDEIKQÖIDKLKKVPNKTMENE-DPKE	798
QY	796	LEKELDLLTDLVEGKGPQAÖTWOVGYLLKTPLPLEPEYIIGLVNY--FDKSGKLIYALD	853
Db	799	IEKKIENIVEKIDKK-----KNIYKEIDKLNEISKIE	831
QY	854	MSDPTIGEÖQKD--AYGNPILNV-----DEDNEGYHTIAV--ATLADYEGEYIK-----	897
Db	832	NDKTSLEKLKNINLSYKSLGNLFLQÖIDEKKKAHEHTIKAMEAYIDDLDNIRKKSQÖIE	891
QY	898	-----DI-----LNSLDKIKAIRQIPLAKYHRLGIFQAIRNMAAEADRLLPK	940
Db	892	KEMNINMDIKMDIHKEMKALNISHDYKIYH--TTSKNHEEKISDIRKNSL---KIÖD	945
QY	941	-TPKCYLINEVPNYRKQKÖMEKN-LKEVDYKTPI-----FNKALPNEKY	980
Db	946	FSEESYINDI---KKELEKQVLESQÖNNNTDINOYLSKIENIYNILKLNKIKKILDKVKE	1001
QY	981	DGDBRAKGH-NINAETNNS--VAVTPIRSEQÖLHKSQSDVNLPÖTSSKNNFYIELIGVSL	1038
Db	1002	YTDELEKNNKKINAELSNSEKIIÖLKENSSLKECQÖSKI---KSTIDDNVVSECTKNITN	1058

QY 1039 CLLEFLV 1044
Db 1059 LKTYIV 1064

RESULT 9

A90551
conserved hypothetical protein MYPV_3130 [imported] - Mycoplasma pulmonis (strain UAB CT
C/Species: Mycoplasma pulmonis
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C/Accession: A90551
R;Cham baud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A/Reference number: A99512; MUID:21267165; PMID:11353084
A/Accession: A90551
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2819 <KUR>
A/Cross-references: GB:AL445566; PID:gl4089727; PIDN:CAC13486.1; GSPDB:GN00153
A/Experimental source: strain UAB CTIP
C/Genetics:
A/Gene: MYPV_3130
A/Genetic code: SGC3

Query Match 4.0%; Score 212.5; DB 2; Length 2819;
Best Local Similarity 18.2%; Pred. No. 0.12;
Matches 246; Conservative 194; Mismatches 444; Indels 467; Gaps 56;

QY 1 MTKKHLKTLALA---LTIVSVTVYSGEVYGLERESV-----KQEQTSASEDDWFEED 51
Db 3 MKNKILIGLGLASWAGITTSAILLPK--IGKGDVAVVNDKNKNDSETSSGSVSDDFDKN 60
QY 52 NEKRTNVSKENSTVDETVSDLFSDGNSNMSSTESVSDPKQVPKAKPEVTQASNSN 111
Db 61 TKHSPDDKANQ-----NKDKSSQSRSESD-----LENKNSEN 96
QY 112 DASKVEVPKODTASKKETLETSTWEAKD---FVTRGDTL-VGF-----SKS 153
Db 97 NFDSGFIDNNSTLSNDKRL-----IKDKIRFVTFGDSIAAGFNAKMLDYVPGEYDENTK 150
QY 154 GINKLSQTSHLVLPASHADGTLQTVASFAFTPPDKTAIAEYTSRLGE-----NGKPSR 207
Db 151 QITGLSYSSYIADYINDLPNKLESFKNFAFS--GTTLKDWNDYFNKNDYFENSTKSE 207
QY 208 LDIDQKEIIDEGEIFNAVQTLKLTIPNGYKSIGQDAFVNDKNIAEVLNPESELTISDYAF 267
Db 208 -DLSQREDF-----LKRLT-----NSNLLLNLTGAN-----DF 234
QY 268 AHMSLKQVKLPDNLKVTIGELAFEDNQIGKLYLP-----RHLLI-----KLAEBAFKSNRI 317
Db 235 MSLAFKFKKDVLSLMKKDASTNDLIG-VFLPLISSIRHEMGIRYKELISHIRKYNKN 293
QY 318 QIVEFLGSKL-----KVIGESAFQDN-----NLKNVMLPDGLEKIESEAFGTG 359
Db 294 VTINLIQYMPALKLFNNVKNLIGESVKIGNDTLGLFLNFINSSIKDQVDLYENVNPN 353
QY 360 NPGD---EHN-----NOVLRTRTGQNPQOLATE-NTYVNPDK-SLMRAT- 400
Db 354 AYDDPLMLKHNDFTDVAFDIHNEIGYKQMAQEIFLKLALDISKYNPSEINSSWNANY 413
QY 401 EDMDYTKWLE-EDFTYOKNS-----VTGFSNKGLOKVRNKNLEIPKQHNGITTEIGDN 454
Db 414 LSYDHGKFSRLFEFKNKDSEIITTVLGVNNSALFSKNDKRIALYESKFTNDYSEI--- 470
QY 455 AFRNVDFOSKTLRKVDLEETIKLPSTIRKIGAFAFQSNULKSFEASEDLBEIKEGAFMNNR 514
Db 471 LFRYSDI-FKTLSSAVE-----FTNSSIFKTLDPEDL---KKFLFANNE 513
QY 515 IGTLLDKDLIKIGDAFHNIHIVAVLPESVQEIGRSAFRONGALHLMFGNKYKTIGE 574
Db 514 ANYQKLFEILTKSRFVHNINRSIRDTLTLLDNNNGAQIGQIEILSHLMTIFKK-----E 568

QY 575 MAFLSNKL-----ESVNL-----EQOKTIEVQAFSDNALS 607
Db 569 ISNLDNYLGLLREHSSSSFYVDPETNLSEFKDIKKILKNFLQKSELKTLISKFTGAFIQ 628
QY 608 EVLPENLOTIREAEFRNHLKEVKGSSSTLSQITFNAPDQNDGDKRFGKVVYV----- 661
Db 629 NKILSHNLSEL-----VLRIVSGEKTFALI-----EKLIDIDFNKAVQYQVEKIE 674
QY 662 -----THNNSHMLADGERFIID-----PD----- 680
Db 675 DELFAFVNTNEKATDFVRFVLNELLVAKPENFKALVSFGISQFLNSSNEVISTQTSQAFY 734
QY 681 KLSSTMYD-----LEKVLK-----IIE-----G 698
Db 735 EVSKTLVNIVDTEELLEKVLKNFIEETKSSAQKLKSNPKENLIEKLQKSIISITFTND 794
QY 699 LDYSTLRQTTQT-----QPREMTAGKALLSKSNLRQGEKQKFLQEAQFFLGRVDLDK 751
Db 795 LGWSFLKQVVKSSKSLSSVEQORDKFTSSLKEVVSFVFKSENVKNKLISSAGTYISNIELEK 854
QY 752 AIAKAKALVTK--KATKNGHLERSIN-----KAVLAYN-----NSAIIKA 791
Db 855 ENSDVPKGFISHINFEKTPPELLNKIIDYVINSWESIDQYNDAFEFTVNLPLKTNSSWEKT 914
QY 792 NVKRLKELDLTDLVEGKPLAQTMTGQVYLLKTPPLPEYYIGLVYFDKSGKLI-- 849
Db 915 YVK-----DIFSSVNSEIOYA-----NLTSKINKLASQ 944
QY 850 YALDMSDTIGEGQDAYGNPILNVDENEGHTLAVALADYEGLYIKIILNSSLDK--- 906
Db 945 AGFELTNSSVESIKNIFNTLKIADSN-----TIDALVD--TFVNSKTLLEIDKNE 995
QY 907 ---IKAIRQIPLAKYHRLGIFQAIRNAAAEADR----- 936
Db 996 IASLKNIQIFALSDDNFVSFKALISSNKLKDKAFVNOYKODITNVKLKELLEKEFENLI 1055
QY 937 ---LLPTPKGY-----LNEVPNRYKKQMEKNLKP----- 963
Db 1056 YSEMPKTSISEFPENPEALNKTTTLRNALKKNDPDPFINKMDIIFDNLDKQKLENYG 1115
QY 964 -----VDKTPIEFKALPNEK-----VDGDRAAKGHNINAEATNNSVA 1000
Db 1116 QVLGLIKDNSDAIATKVQAIKFDVNVNEANQDVLKVLVSTKKFMGEFELNADETNIKIS 1175
QY 1001 VTPIRSEQQLHKSQSDVNLPTQSSKNFIYE 1031
Db 1176 TDLTKLKQVFESNNFAKSLALATKDSLVID 1206

RESULT 10

T39068
coiled coil protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T39068
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Connor, R.E.; Wood, V.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z21825
A/Accession: T39068
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1208 <MUR>
A/Cross-references: EMBL:Z81317; PIDN:CAB03608.1; GSPDB:GN00066; SPDB:SPAC6G9.06c
A/Experimental source: strain 972h-; cosmid c6g9
C/Genetics:
A/Gene: SPDB:SPAC6G9.06c
A/Map position: 1

Query Match 3.9%; Score 209; DB 2; Length 1208;
Best Local Similarity 19.4%; Pred. No. 0.051;
Matches 220; Conservative 170; Mismatches 367; Indels 378; Gaps 52;

```

QY 20 TYSQEVYGLEREBSVKQEQTOSA-----SEDDWFEEDNERKTNVSKENSTVDETVSDFLS 74
Db 93 SFDPNNGYGL---SAISKQATQEAALSISQGNDSY---DVSKLTDLISK-NSEIDHTDGL-- 143
QY 75 DGNNSNSSKTESVSD-----PKQVPEKA---KPEVTQEAASN 108
Db 144 PANAAALTLREQEKVLEKVSRENFGRLIKIVCLEKRLSEMAPEQIKVAVKDNVELHAERAN 203
QY 109 SSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQ-----T 161
Db 204 LQQLKRTESLLQKSEDKNFKE---EKVDYLSK-----VNDVEQSQNVKVF 248
QY 162 SHLVPSHAADGTQLTQVA-SFAFTPPDKTAIAEY-----TSRLGENGKPSRLDIDOK 213
Db 249 ERIRFLENALKEVQREKDSLSTEMEEDKSNKEVDYEYERQLOQRULDELSE--ELDVAQ- 305
QY 214 EIIDEGEIFNAYQLTCLKLTPNGYKSGIDAFVDNKNIAEVLNLP----- 257
Db 306 DLITEKE-----DEIATLKROIEEKENSSSAFENEENSSYVHLQEDYAILQAKCDEFADRI 361
QY 258 -----SLETSDYAFAHMSLKQVKLPDNLKVIIGELAFDNOIGKLYLPRHLIKLAERAF 312
Db 362 QVLTADLEKEKENQIMHESEASIGLTDSMQY-----H-TLQEQHL 400
QY 313 KSNRIQTVFEGSKLVIGE--ASFQDNNLRNVMLPDGLEKIESEAFNGPDEHYNNQV 370
Db 401 KAN--EEIEFLHDQISRMNEEGKNFEDIMLOFRSLBERDVLSEKLOT---LEDNNSL 454
QY 371 VLRTRT-GQNPQOLATENTYVNPDKSLMKATPDMDYTKLBEEDFYQKNSVTGFSNKGLO 429
Db 455 RLMTSSLGNQIESLRTQNRIDEDEKNHLRLASKNSDKALAE-----TNIRLO 502
QY 430 KVRRNKNLEIPQONGITITEIGNAFRN-----VDFOSKTLRKYDLE 472
Db 503 EV--TKELETLRMKNSNDLNEIHLRENEGLTLKIDSITKEKDRILNELEQRIKSYEVN 560
QY 473 EIKLPSTIRK-----IGAFAFQSNLKSFEASEDLKEIKEGAFMNNRIGT 517
Db 561 VSELNGTIDEYRNKLKDKETYNENVNAFOYKXNDLRFHESINKLQOREKELTSN---- 616
QY 518 LDLKDKLIKIGDAFHINHIYAVLPESVQEIGRSAFRONGALHLMFIGNKYKTIGEMAF 577
Db 617 LEKKNLVI-----SSLRETVAMLEKERESIKKYLSGN----- 648
QY 578 LSNKLESVNLSE-----QKOLKTI--EVOAFSDNALSEVVLPPNLQITREAFKRN 626
Db 649 -AKDLDTNLMETLNDKISVLOQOLTDVKDELVDSEEREBAIVAGOKLASAFELMSNEK 707
QY 627 HLKEVKGSSSTLSQITFNAFDONDGDKR-----FGKKVV---VRTHNNSHMLADGERFI 676
Db 708 QALELKYSLSKNEL-INA--QNLDRREBELSELSSKLLPEERKIRSGSNDIEKNKEINV 764
QY 677 ID---PDKLS-----STWVDLEKV-----LKIEGLDYSTLR- 705
Db 765 INSELADKLAQIRHLESDBKMLDKLVHHLNRGIBERANIBENAVKRLCLMGCDDYSSVSI 824
QY 706 -----QTQOQFREM-----TTAGKALLSK 725
Db 825 LQIVSQIEHFVNQIQITRSLKQELRHDVVOFGSKKEQELSRSFEKFGLTETKHDILAQ 884
QY 726 SNLRQGEKQKFLQE-AQFELGRVD-----LDKAIKAKEKALVTYKATK 767
Db 885 RNRNVSEKMDLENAQAQFSSPDRKNGYLYPSEHTSKIEYLEKTIEDLKAL--QDELK 942
QY 768 NGHLLERSINKAVLAVNNSAIK-KANVKRLKEKELDLTD-----LVEGKG 811
Db 943 NRNLMDDIS---SYNKQTTKLQEKIKWLERERSILIDELBSYRSNOFNQNNLVQDKN 998
QY 812 PLAQATMVQGVYLLKTPLEPYIIGLVN---FDKSGKLIYA--LDMSDTIGEGQDAY 866
Db 999 ELEER-----LKEIQELELVYNNHFMKQAEIMTSNVTDESQMLKTLREAL 1044
QY 867 GNPILNVEDNEGHTTAAVATLADYEGLYIKDILNSSLDKIKAIRQIPLAKYHRL 921

```

```

Db 1045 QSKTNNDH---LSTLIERNRKKEYKSL-----LDDYNQLR---ARYKNL 1082
RESULT 11
S38173
myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YKR095w; protein YKR415
C/Species: Saccharomyces cerevisiae
C/Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 21-Jul-2000
C/Accession: S38173; S40647; S31207
R/Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo
submitted to the Protein Sequence Database, March 1994
A/Reference number: S38158
A/Accession: S38173
A/Molecule type: DNA
A/Residues: 1-1875 <BAL>
A/Cross-references: EMBL:Z28320; NID:g486586; PID:g486587; MIPS:YKR095w
A/Experimental source: strain S288C
R/Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.;
Yeast 9, 1349-1354, 1993
A/Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chro
A/Reference number: S40644; MUID:94205265; PMID:8154186
A/Accession: S40647
A/Molecule type: DNA
A/Residues: 1-1875 <BOU>
A/Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A/Experimental source: strain S288C
R/Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A/Title: A new yeast gene with a myosin-like heptad repeat structure.
A/Reference number: S31207; MUID:93247549; PMID:8483450
A/Accession: S31207
A/Molecule type: DNA
A/Residues: 1-300, 'A', 302-1875 <KOE>
A/Cross-references: EMBL:L01992; NID:g171958; PIDN:AAA34783.1; PID:g171959
C/Genetics:
A/Gene: SGD:MLP1
A/Cross-references: SGD:S0001803; MIPS:YKR095w
A/Map position: 11R
Query Match 3.9%; Score 209; DB 2; Length 1875;
Best local Similarity 19.2%; Pred. No. 0.098;
Matches 221; Conservative 194; Mismatches 375; Indels 362; Gaps 55;
QY 32 ESVKQEQTQASAE-----DWEFEEDNERKTNVSKENSTVDETVSDL-----FSDGNSN 79
Db 107 ESVENKMKJSSLEFVKRKLDLITEKKEQTQSNQORTLILDERIKEIELVRVENNRSN 166
QY 80 NSSKTESVSDPKQVPKAKPEVTQEAASNSNDA-SKVEVPKQDTASKKETL--ETSTWE 136
Db 167 SECKKLRTIMDL-----TKQGYTINDLNSRTLELRK--TQELTLLQSNNDWL 214
QY 137 AKDFVTRGPTLVGFS-----KSGINKL-----SQTSHLVPSHAADGTQLTQV- 179
Db 215 EKELRSKNEGYLSYRQKTDKVIDLRNEINRLRNDQMERITNDVVKQKNNELSKSLQEK 274
QY 180 -----ASFATPPDK-----TATAEYTSRLGENGKPSRLDIDOK 213
Db 275 LLEIKGLSDLSNSEKQEFSAEMSLKQRLVDLLESQNLNAVKEELINSIRELNTAKVADDSK 334
QY 214 EIIDEGE-IFNAYQLTCLKLTPNGYK-----SIGODAFVDNKNIAEVLNLPESLETISDYA 266
Db 335 KQTPENEDLKEIQLTKEKLAQCEKECLRLSSITDEADEDNENL-----SAKSSSDF- 386
QY 267 FAHMSLKQVKLPDNLKVIIGELAFDNOIGKLYLPRHLIKLAERAFK---SNRIQT-VEF 322
Db 387 -----IFLKQOLIK--ERRTKEHLQNOIETFIVE 413
QY 323 LGSKLKVIIGBASFOQNNLRNVMLPDGLEKIESEAFNGPDEHYNNQ--VVLRTRTGQNP 380
Db 414 LEHKVPIIN--SFKERT-----DMLENELNNAALLLEHTSNEKNAKVKELNAKNO 461

```


QY 381 HQLATENTVNPDKSLWRATPMDYTKWLEEDFTYQKNSVTGFSNKGLOK--VRNNKLE 438
Db 462 KLVECENDLQTLTKQ-----RLDLCRQIQ--YLLITNSVNSDKGPLRKEEIQIONI- 512
QY 439 IPKQHNGITTEIGDNF--RNVDFQSKTLRKDYLEIPLSTIRKIGAFQFQSNLKS 495
Db 513 --MQEDSTITESDQKVTERLYEF--KNIIQLOEKNAELKVVRNL-ADKLESKEKKS 567
QY 496 FEA-----SEDLIEKEGAFMNNRIGTLDDKLIKIGDAFHINHIVLPSVQIEG 550
Db 568 KQGLQKIESETVNEAKE-ALITLSEKMDLESRIE-----LQKELEL- 610
QY 551 RSAFRONGALHLMFTGNKVTIGEMAFLSNKLBSVNLSEQOKLTIEVQAFSDNALSEVV 610
Db 611 -----KTSVNEEDASYSN-VTIKQLTETKRDLESQVQ----- 641
QY 611 LPPNLQITREAFKRNHLKEVKGSSSTLSQITFNAFDQNDGDKRFGKVVVRTHNNSHMLA 670
Db 642 ---DLQI-----RISQITRESTENMSLNKEIQDLYDSKDISIKLGKEKSSRILA. 689
QY 671 DGERFIIDPKLSSTWVLEKYLKIEGLDYSTLRQTTQI--FREMTTAGKALLS----- 724
Db 690 E-ERFKLSNTLIDLTKAENDQLRKRPDYLONTILKQDSKTHEFTINEYVSC-KSKLSIVET 747
QY 725 -----KSNLR-----QGEKQFLQEAQFELGR 746
Db 748 ELNLKKEEQKLRVHLKRNKLQELNKLSPKEDSLRIMVTQIQTQKEREDLLEET-----R 802
QY 747 VDLKALAKAEKAL-VTKKAT--KNGHL--LERSINKAVLAYNNSAIKKANVKRLKEELD 801
Db 803 KSCQKIDLELDALSELKETSQKDHIIKQLEBDNNSNIENYQN-----KIEALKQYE 856
QY 802 LITDVEGKPLQATMVGCVYLLKTPPLPEYIIGLVNVPDKSGKLIALDMSDTIGEG 861
Db 857 SVITSVDSK---QTDIEKLQYKYS-----LEK----- 881
QY 862 QKDAYGNPILNVEDNEGYHTLAV--ATLADYEGLYIKDILNSSLDKIK-----AIRQIP 914
Db 882 -----EIEEDKIRLHTYVMDDETIND-----DSLRLKELEKSKINLTDAYSQIK 924
QY 915 LAKYHRLGIFQAIRNAAEADRLLPKTPKGYLVNVPYRKQOMEKNLKPVDYKTPLEN-- 972
Db 925 EYK---DLYETTSQSLQQTNSKLDESFPDFTNQIKNLTDEKTSLEDKISLKEQOMENLN 980
QY 973 -----KALPNEKVDGDRAAKGHINAETNNSVAVTPIRS--EQQLHKSQSDVN--LP 1020
Db 981 NELDLQKGMKEKADFKK-----RISILQNNKKEVAVKSEYSEKLSKIQNLDLQQTIV 1035
QY 1021 QTSKKNFIYEI 1032
Db 1036 ANTAQNMYEQEL 1047

RESULT 12
B89921
hypoetical protein ebha [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: B89921
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
L:ncet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: B89921
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6713 <KUR>
A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BAB42527.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ebha

Query Match 3.9%; Score 209; DB 2; Length 6713;
Best Local Similarity 17.6%; Pred. No. 0.65;
Matches 199; Conservative 207; Mismatches 406; Indels 318; Gaps 46;
QY 29 ERESVKQEQTQASQEDDMFEEDNERKTVSKENSTVDETVSDLFSDGNSNNSKTE-- 86
Db 5190 DKKEAVDQALQAAQSTIDFTNGSNANKDAVEQALTKLQEKVNL--NGNERVAEAKTQAK 5247
QY 87 -----SVVSDPKQVPKAP--EVTQASNSNSNDASKVEVPKQDTASKKET 129
Db 5248 QITDQTLNADQIATAKONIDQATKLQPIAELVDQATQLNQSMQDQQAQVNEHANEQIT 5307
QY 130 LETSTWEA-----KDFVTRGDTLVGFSKSGINK--LSQTSHLVLP SHAADGTQLTQVAS 181
Db 5308 IDVTQADSDKQAKYQQAIDAENVL--KQNAKQQVQALQNTLN-----AK 5352
QY 182 FAFTPDKCTAIAEYTSRLGENGKPSRLDIDQ-----KEIT 216
Db 5353 QALNGDERVALAK-----TNGKH--DIDQLNALNNAQDQGFKGRIDQSDNDLQIQIIV 5403
QY 217 DEGEIFNAY--QLTKLTIPIGKYSIGQDAFVDN-----KNIAEVLNPLESETISDYAFAM 270
Db 5404 DEAKALNRAMDQLSQETIGNEGRTKSTNYVNAQTQYQYDEAVDQAKQALDKSSGQNL 5463
QY 271 SLKQV-KLPDNL-----KVIGELAFPDNQIGKLYLPRHLIKL--ABRAFKSNRIQIVFEL 323
Db 5464 TAEQVIKLNDAVTAAKKALNGEERLNRRKAELQRLDQLTHLNNAQROLAIQIINNAETL 5523
QY 324 GSKLVIGEASFQDNNLRNVMPLDGLKIESEAFTEGDEHHYNNQVLRTRTGQNPQL 383
Db 5524 NKASRAINRATKLDNAM-----GAVQYIDEQHLGVISSTNYIN----- 5562
QY 384 ATEENTVNPDKSLWRATPMD-----YTKWLEEDFTYQKNSVTGFSNKGLOYRR 433
Db 5563 ADDNLKANYDNAIANAAHELDKVGQNAIAKAEAEQLKQNIIDAQNALNGQNLANAKDA 5622
QY 434 N-----KNLEIPKQHNGITTEIGDNFANRVDFQSKTLRKDYLEIPLSTI 480
Db 5623 NAFVNSLNGLNQOQDLAHKAINNADTVSDVTDIANNQIDLNDAMETLKLHLDVNEIPNAE 5682
QY 481 RKIGAFQFQSNLKSFEASEDLIEKEGAFMNNRIGTLDDKLIKIGDAFHINHIVAI 540
Db 5683 QTVNYQNADNKTNF--DDAKRLANTLLNSDNTNVNDINGALQAVNDAIHNLNGDQRL 5739
QY 541 --VLPESVQIEGRSAFRONGALHLMFTGNKVTIGEMAFLSNKLBSVNLSEQOKL--K 594
Db 5740 QDAKDKAISINQA-----LANKLK-----EIEASNATDQDKLIAKNK 5777
QY 595 TIEV-----QAFSDNALSEVLLPPLQITREAFKRNHLKEVKGSSSTLSQITFNAF 645
Db 5778 AEELANSIINNINKATSNOAVSQV-----QTAGNHAIQVHANEIPKAKI----- 5822
QY 646 QDNDGDKRFGKVVVRTHNNSHMLADGERFIIDPKLSSTWVLEK-VLKIEGLDYSTL 704
Db 5823 --DANKVDVKQV-----QALIDEIDRNPNLTDKEQALK-----DRI 5857
QY 705 RQTTQIQUFEREMTTAGKALLSKNSLRQGEKO--KFLQEAQFPL-GRVLDKALAKAEKALV 761
Db 5858 NQIIQGHNDINNA---LTKEIEQAKAQLAQLAQDIDKIDLVAKEDAKQDQVQAL 5913
QY 762 TK-----KATKNGHLERSINKAVLAYNNSAIKKANVKRLKEKLDLLTDLVEGKPLAQ 815
Db 5914 DEIDQNPNTLDKEKQALKDRINQ--ILQGHNGINNAMTKE-----EIEQAKAQLAQ 5963
QY 816 ATMWQGVYLLKTPPLPEYIIGLVNVPDKSGK-LIYALDMSDTIGEGQKDAYGNPILNV 874
Db 5964 A-----LKEIKDLVKAKENAKQDVQVQALIDEIDQNPNTLDKEKQALKDRINQIL 6015
QY 875 EDNEGHTL-----AVATLADYEGLY-IKDILNSSLDKIKAIRQIPLAKYHRLG 922
Db 6016 Q--QGHNDINNMTKEIEQAKAQLA--QALQIDIKDLVKAKEDAKNAIKALANAK----- 6066

QY 923 IFQAIRNAAEADRLLPKTPKGYLINEVPNRYKKQMEKNLKPVDYKTPIFNKALPNEKVDG 982
Db 6067 -----RDQINSNPDLTPEQKAKALKEID-----EA 6091
QY 983 DRAKGNINAEITNSVAVPIRSEQOLHKSQSDVNLPTSSKNFIYEI 1032
Db 6092 EKRALQNVENAQTID-----QLNRG---LNLGLDDIRNTHWMEV 6127

RESULT 13

C97033
uncharacterized protein, probably surface-located [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: C97033
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97033
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1227 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK79054.1; PID:g150233995; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1080

Query Match 3.9%; Score 208.5; DB 2; Length 1227;
Best Local Similarity 18.9%; Pred. No. 0.055;
Matches 221; Conservative 164; Mismatches 363; Indels 419; Gaps 54;

QY 3 KKHLLTALALTTVSVTYSGEVYGLERESVKOEQTQASG-DWFEEDNERKTNYSKE 61
Db 5 KNILLVFSISVTSVAVLLSKPAIKAYAADNSVYKSSISNSNEINVKKEVQKEVNVNKN 64
QY 62 NSTVDETVSDLFSDGNSNSSSKTESVSDPKQVPKAKPEVTQASNS-----NDAS 114
Db 65 N-----DSNSKVSSSNENGVSNKENSNPVSSSSSEIQSINKVNLQVQNNKS 112
QY 115 KVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLP SHAADGT 174
Db 113 VLAASNVDDEVK--INSGNVQTSYIAIGETKV----- 144
QY 175 QLTQVASFATPDCKTAIAEYTSRLGENKPSRLDIDQKEIID-----EGEIFNA 224
Db 145 -----KPDITDIINKAIVDARSABGTDLSEVEIYDI 175
QY 225 YQTLKLTIPNGYKISGQDAFVNDKNIAEVLDESLETISDYAFAMSLKQVKLPDNLKVI 284
Db 176 VSQTAELI-----QFAF-----RNINDGVANVSDYT-----LL 203
QY 285 GEIAFFDNQIGKLYLPRHLIKLAERAFKSNRIQTVEFLGSKLVIGASFOQNNLRYVM 344
Db 204 G--ATFVNDAN-----LDSVNKYFYHKKRYATV---TKFKDV--ATKTSNALKNI- 245
QY 345 LPDGLKIESEAFITGNPGEDEHYNNQVVLRTTRGQNPQALATENTYVNPDKSLWRATPDM 404
Db 246 -----NN-----GQG-----GETD 254
QY 405 YTKMLBEDFTYQKNVTFGSNKGLOKVRNKNLEIPKQNG--ITITEIGDN----- 454
Db 255 YTA-LE-----VSGVTQPYLDLV--KNKIYKEKQNGRDLTITEISDASAATIARI 302
QY 455 --AFRNVDFQSKTLRY-----DLEIKL-----PSTIRKIGAFQSN 491
Db 303 NTALDNMDAGVAITLEDYQAI GANNVPQLHVADVNSLAMDQKWDVSEALDGINTIMTYIN 362
QY 492 NLKSFASEDLLEEIKGAFMNNRIGTLD-----LKDKLIKIGDAAFHNIYALVP 543
Db 363 NINSGVGTED-DYINSHAVDSNE-GNIDYDILNANIIEKKTAKGQD-----LTIP 410

QY 544 E---SVQEI-----GRSAFR--QNGALHMFIGNKVKTIGEM-----AF 577
Db 411 EVANVVKVETKLLIDFYNHAAAGQTTLDQYKAVDPNAQVOQDDVATLSMLKTRDCKTLKA 470
QY 578 LSNKLESV-----NLSEQKLTIEVQAFSDNALSEVVLPPNLQITREBAFKR 625
Db 471 LQDKIDSLNSLKNINSIGINIDYSKIQTEAVDASKLEAVND-----DIKKIKADKGRD 525
QY 626 NHLKRVKGS--STLSQITFNAFDQNDGDKRFKGVVVRTHNNSHMLADGERFIIDPKLS 683
Db 526 LTIQIRDSVAKTIDYIN-STSNVSKGDSVSDYITIGIDVTEINIE--FVNERIKES 581
QY 684 STMVDLEKVLKIEGLDYSLRQTTQTQFREMTTAGKAL-----LSKSNLR----- 729
Db 582 GITTIENIKVIE-----PIVQSEVYVRITGVGVVDYKTLGINNVNDNNIYINAE 636
QY 730 -QGEKQKFLQEAQF-----FLGRVLDKAIKAEKAL-----VTKKATKNGHLLERSIN 777
Db 637 LKNKKDVKIQDIQTRVDNTINNIDVINKIGAGDAVLSDYFNIGIT-----DYY 684
QY 778 KAVLAYNNSAIKKANVKRLKEKELDLTDLVEGKGLAQATMVQGVYLLKTPLPDPEYIIG 837
Db 685 QDILDYVNAADLKIQNY---KDVDIIEERVEAKISSYEALM-----RINIGEAVTDDFKA 735
QY 838 LNVYFDKSGKLIYALDMSDTIGEGQDAYGNPILNVEDNEGYH-----TLAVATLADYE 892
Db 736 LGITDINDGLLYA--TTDLQNKNYKTA-DEVIAVQAQIEIRYALMQINLGKATTADYN 792
QY 893 GLVIKDIINSLDKAIKAIKQIPLAKYHRLGIFQAIRNAAEADRLLPKTPKGYLN--EVP 950
Db 793 TLGIIDVNDISILTYVNAADIQ-----GKNYVNVDEV- 822
QY 951 NYRKKQMEKNLKPVDYKTPIFNKALPNEKVDGDRAA---KGNINAETNSVAVTPIRS 1006
Db 823 ---KAEIEKNIQ-----IYNALL--KIDSGSATIDYRTIGITTVIDYINISYVNIIRI 869
QY 1007 EQOLHKSQSDVNLPTQSSKNFIYEIL 1033
Db 870 K-----GMNIPQVSDAKRFINIIIL 888

RESULT 14

B70126
surface-located membrane protein 1 (Imp1) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 11-Jan-2000
C;Accession: B70126
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Iathigra, R.; Whi
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vug
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: B70126
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1119 <KLE>
A;Cross-references: GB:AE001131; GB:AE000783; NID:g2688098; PIDN:AAAC66595.1; PID:g2688
A;Experimental source: strain B31
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat
F;742-774/Domain: tetratricopeptide repeat homology #status atypical <TT1>
F;775-808/Domain: tetratricopeptide repeat homology <TT2>
F;809-842/Domain: tetratricopeptide repeat homology <TT3>
F;843-876/Domain: tetratricopeptide repeat homology <TT4>
F;911-943/Domain: tetratricopeptide repeat homology <TT5>
F;944-977/Domain: tetratricopeptide repeat homology <TT6>
F;979-1010/Domain: tetratricopeptide repeat homology #status atypical <TT7>
F;1011-1044/Domain: tetratricopeptide repeat homology <TT8>

Query Match 3.9%; Score 207.5; DB 2; Length 1119;
Best Local Similarity 19.4%; Pred. No. 0.053;
Matches 223; Conservative 184; Mismatches 399; Indels 341; Gaps 59;

```
QY 35 KOEQTQASSED-DWFEEDNERKTNSKENVSTV-----DETUSDLFSDGNS----- 78
Db 94 KDLQDSANQILDKIKGQDNTKTWN-ENPDIAFNNRYIKDSTITTENTYSDRNDVIGIDE 152
QY 79 -----NNSKTESVSDPK-----QVPKAKEPTV--QEASN 108
Db 153 DISEFKKSKIPEKIKPNTNPKXEDQIIQSPNPKLSVNDQKNLFNLEKTKKNLSGKSNSEN 212
QY 109 SSNDASKVEVPKQDITASKETLETSTWEAKD---FVTRGDTLVGFSKSGINKLS---QT 161
Db 213 ILNDSQKIENDKQNTNLSKSKNSENILKTPDNSKYSNNNT-----TSLKKSISNSQKE 266
QY 162 SHLVPSHADG-----TQLTQVASFATPDKKTALIAEYTSRLGNGKPSRLDIDQKEI 215
Db 267 SELSPPSQTIIGKIYRPSYSLIKKELYRILDDINTG---RVTLGKN---RL-----KEL 314
QY 216 IDEGEIFNAYQ-----LTUKLTENGYSKISGQDAFVNDKNIAEVN- 254
Db 315 IKKG-LSNKFQKVNELIENSKNKEASNLTLTIKKDIEPNLINIPKDPY--KKEIFQLDK 371
QY 255 ---LPESLETISDYAFAMSLKQVLPDN-----LKVIIGELAFPDNOIGKLYIPRHLI 305
Db 372 EDKPPQYLEDLKSQVH--SIKPIDLENTKSRQQAIDKQNE--FLKNN-----PND-A 418
QY 306 KLABRAFKSNRIQTVEFLGSKLVIGEASFQDNNLRNVMIPDLEKIESEAFNGDDEH 365
Db 419 QASKTLAQANKIQHLEDLKSQVHSIKPIDLENTKSRQQAIDK----LNEFLKNNPNDQAQ 473
QY 366 YNNQVVLRTRGQNPHQIATENTYVNP-----DKSLWRATPMDYTKWLEEDFTYQKNSV 420
Db 474 ASKTLAQANKI-QHLEDLKSQVHSIKPIDLENTKSRQQAIDKLN-----EF-LKNNPN 524
QY 421 TGRSNKGLQKVRNKNLEIPKQHNGITTEIGDNAFRNVDFOQK--TLRKVDLEIK-LP 477
Db 525 DAQASKTLAQANKIQHLE-----DLKSQVHSIKPIDLENTKSRQ 563
QY 478 STIRKIGAFAFQSNNLKSFEASEDLLEIKEGAFMNNRIGTL-DLKDKLIKIGDAFHNH 536
Db 564 QAIKDINEFL--KNPNDAQASKTLAQ-----NKIQHLEDLKSQ-----VHS 604
QY 537 IYAIVL-----PEVQEIERSAFRONG-----ALHMFIGNKVKITIGEMAFLSNKLSEV 585
Db 605 IKPIDLENTKSRQQAIDKQNEFLKNNPNDQAASKTLAQANKIQHLEDLKSQVHSIKPI 662
QY 586 NISEQKQKLTIEVQAFSDNALSEVVL--PPNLQOTIR--EBAFKRNHKEVKG----- 633
Db 663 DLENTKSRQ---QAIKD--LNEFLKNNPNDQAASKTLAQANKIQHLEDLKSQVHSIKPI 716
QY 634 --SSTLSQ---ITFNAFDQNDGDKRFQKVVVVRTHNNSHMLADGERFIIDPDKLSSTMV 687
Db 717 DLENTKSRQQAIDKQNEFLKNNPNDQAASKTLAQAYENNNDLLKAENAYEKIKLNTQOE 776
QY 688 D-----LEKVLKITEGLDYSTLRQTTQTOFRMTAGKAL--LSKSNLRQGEKOK 735
Db 777 DHYKLGILIRFKLKKEVHESIESFD-QTIKLDPKHK-KALHNKGIALMLNKNKKAIESREK 834
QY 736 FLOEAQFELGRVDLDKALAKAEKALVTKK-----ATKNGHLLERSINKAV----- 780
Db 835 AIQ-----IDKNYGTAYIQKGIAEKNKGMQAFASFQNAYNLDKNPNYALKAGIVS 886
QY 781 -----LAYNSAIKKANVKRLEKBLDILTDLVE----- 808
Db 887 NNIGNFKQSEYYLNFNANAKKPEIAIYNLSIAKFENNKLSESLETINKAIDLNPEKSE 946
QY 809 -----GKGPLAQATMVOGYLLKTPPLPEYYIGLVNYFDKSGKLIYALD-MSDT 857
Db 947 YLYLKASINLKKENYQNAISLYSIVIEKNP-ENTSAYINLAKAYEKSQNAISTLEKI 1005
QY 858 IGEQKDAYGN-----PILNVEDNEGYHTLAVATLADYEGIYIK 897
Db 1006 INKNMIALNNLIGILYKKEKQYQKAIIEFEKAIIN--SDIEAKYNLATTLIEINDNTRAK 1063
```

```
QY 898 DILNSLDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPKTPKGYLNEVBNYRKQOM 957
Db 1064 DLLR-EYTKLKPNNEAL---HALGIIEYNNENNNDQTLREL-----IKKFPNYKNE- 1111
QY 958 EKNLKPV 964
Db 1112 --NIKKI 1116

RESULT 15
T28677
rhoptry protein - Plasmodium yoelii
C/Species: Plasmodium yoelii
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C/Accession: T28677; C45521
R/Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994
A/Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
A/Reference number: Z20508; MUID:95021522; PMID:7935623
A/Accession: T28677
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-2269 <KEE>
A/Cross-references: EMBL:L27638; NID:g457145; PID:g457146; PIDN:AAA21304.1
R/Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A/Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple
A/Reference number: A45521; MUID:91101660; PMID:2270106
A/Accession: C45521
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 2131-2269 <KE2>
A/Cross-references: GB:M34283

Query Match 3.9%; Score 207; DB 2; Length 2269;
Best Local Similarity 18.8%; Pred. No. 0.16;
Matches 226; Conservative 206; Mismatches 432; Indels 340; Gaps 58;

QY 29 ERESVKQEQTQASGEDDWFEEDNERKTNSKENVSTV---ETVSDLFSDGNSNNSSKT 85
Db 989 DHDYNIKELKSHDKSNGYKTEADQNKKAIQKNEHFEQYKEEVTVLNKKYAVELKNKF 1048
QY 86 ESVSDPKQVPK---AKPEVTQEAASNSNDASKV-----EVPKQDTASKK----- 127
Db 1049 DRTKNDKQIIEIKDAHNYCTLESQSEKKMNEIKNEKIHIEDEVANNDKSNKAITSIK 1108
QY 128 ---ETLETSTWEAKDFYTRGDTLVGFSKSGINKLSQTSHLVLPASHADGTQLTQVAFRAF 184
Db 1109 VSEVPFKTIKIKINEIRTKSD-----DCLKETNDLE-----KQISNL-- 1145
QY 185 TPDKKTALIAEYTSRLGNGKPSRLDIDQKEIIDEGEIFNAYQTLKLTIPNGYSIGQDAF 244
Db 1146 -----SIDTQETKLTENGK-----QKLTLE--ELLESLLKQKKNIED----- 1180
QY 245 VDNKNIAEVNLPESLETISDYAFAMSLKQVLPDNLKVIGELAFPDNOIGKLYIPRHL 304
Db 1181 -QKKELDEVN--SKIKNIENTVNOHKKNYEIGIVEKINEIAKTN--KNQIESYKELIKPT 1235
QY 305 IKLABRAFKSNRIQTV---EFLG---SKLKVIGEASFQDNNLRNVMIPDLEKIESEAF 358
Db 1236 IQHITISFNANDLEGIDSDENLGKYNTEMGNIYEFTKSYNL-----ITNYLETYSKESIT 1291
QY 359 GNPGEHYNNOVVLRTRGQNPHQIATENTYVNPDKSLMRATPMDYTKWLEEDFTYQKN 418
Db 1292 -----YNOI-----QN-KRIDQKELKNIEVNNKAKSYLDYIK--ENEF---DR 1330
QY 419 SVTGFSNKGLQKVRNKNLEIPKQHNGI-----TTEIGD-NAFRNVDFOQKTL--- 466
Db 1331 IYTHFKKK-LNTVNDNFKQNEYSKVNEGFDNISINTVKNSTDENSLNLTNGTKEYAN 1389
QY 467 -----RKYDLBEI-----KLPISTIRKIGAFAFQSNNLKSFEASEDLLEIKEGAFMNN 513
Db 1390 IVNNTYYSYKYEABENIPRNIPKLTANTLN-----IKTKNSSGIDLSKDI-KAIALSYDS 1442
```


QY 514 RIGTLDLKDLI-----KIGDAFHINHYAIVLP--ESVOEIGRSAFRONGALHMF- 564
Db 1443 K-----TEDTLIFIPSPQKTEYTKISDSYSILLDLKKSQELQK--KEQQTLLKLF 1494
QY 565 ---IGNKYKIGEMAFLSNKLSESVNSEQOLKTEVQAFSDNALSEVLP-PNLQTI 620
Db 1495 NRRLEYKQATNE---LRGTLSDPKYKKEKILSEVKLLHKSNEKLKSCNFQNYDTLIE 1551
QY 621 EA-----FKRNHLKEVKSSTLSQITENAFDQNDGDKRFGK--KVVVRTNNSHMLADG 672
Db 1552 SSKYDQVKEKSNNYKQEKELGID-----FNVTDMEEKFNNDIKVIELENNYDSSEEN 1605
QY 673 ERFIDPDKLSSTM---VDLEKY-LKIEGLD-----YSTLRQTTQTQ 711
Db 1606 NNILQSKQKLKELTNKFNAEIKIDKIEKNDLIDKLETRKNCMLFTHTTLAETLKI- 1664
QY 712 FREMTTAKALLSKSNLRQGEKQFLQEAQFPLG-----RVDLDKALAKA 756
Db 1665 --KITDYSKFIESAT-----KESKEFLKYIGDTSNSLNDIATLQKYLHQ-INKY 1713
QY 757 EKALVTYKATKNHLLERS-----INKAVLAYNNSAIK 789
Db 1714 VTSKLSDATNDNNNLEKEKEATQAIKNLTKLFTIDSNNIDANALHNKKIQMVYFENSELH 1773
QY 790 KA--NVKRIKELDLTLTDLVEKGPIAQTWQGVYL-----LKTPLPLPEYYIGLV- 840
Db 1774 KSIESIKQLYKKMHVEKFLNIGQ-----INGKYFDISKQFDNILLQOSELTANLND 1825
QY 841 -----YFDKSGKLIYALDMSDTIGEGQDAYGNPILNVEDNEGYHTLA-----V 885
Db 1826 LKEIGQKISDKKNKFLHALNET-----PIPNFTLKEIYHDIKVKYKRQIDEI 1872
QY 886 ATLADYEG---LYIKDILNSSLDKIKAI-----RQIPLAKYHRLGIFQAIRNAAAE 933
Db 1873 ENITSEENENTLTYI-DTITLKEKQVQSLNFTTYENDSNIIKQH---IQDTNENDVSK 1928
QY 934 ADRLLPKTPKGY---LNEVPNYRKKQOMEKN-----LKPVDYKTPIF 971
Db 1929 IKESLKTITQSFQELINKINGIKAQFYDNNNNNNINNTISTISQDVNDYKKHISKDLTIE 1988
QY 972 NKALPNEKVDGPRAKAGHININAEITNSVAVTPIRSEQQLHKSQSDVNLPOTSKNNFIYE 1031
Db 1989 NELIEIQKSLIEDIKNSTYEIRGANNNYVNTIRNYVEQQTNKIQNNSKDEI---DDITQK 2045
QY 1032 ILGY 1035
Db 2046 ILNY 2049

Search completed: April 27, 2004, 11:16:32
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: April 27, 2004, 11:14:07 ; Search time 23 Seconds
(without alignments)
2368.061 Million cell updates/sec

Title: US-10-091-007a-24
Perfect score: 5352
Sequence: 1 MTKKHLKTALALTTVSVT.....VSLCLFLVTAGKKGRARK 1055

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	226.5	4.2	3696	4 US-09-134-001C-5080	Sequence 5080, Ap
2	222	4.1	10186	4 US-09-134-001C-3159	Sequence 3159, Ap
3	201.5	3.8	1786	3 US-08-973-462-8	Sequence 8, Appli
4	200	3.7	872	3 US-08-851-843A-8	Sequence 8, Appli
5	200	3.7	872	3 US-08-851-843A-54	Sequence 54, Appl
6	200	3.7	872	3 US-08-974-549A-221	Sequence 221, App
7	200	3.7	872	3 US-08-854-050-8	Sequence 8, Appli
8	200	3.7	872	3 US-08-854-050-54	Sequence 54, Appli
9	200	3.7	872	4 US-09-430-323-8	Sequence 8, Appli
10	200	3.7	872	4 US-09-430-323-54	Sequence 54, Appl
11	200	3.7	872	4 US-09-402-181B-221	Sequence 221, App
12	200	3.7	872	4 US-09-721-456-221	Sequence 221, App
13	192.5	3.6	2662	4 US-09-595-684B-31	Sequence 31, Appl
14	189	3.5	1221	4 US-09-107-532A-3959	Sequence 3959, Ap
15	188	3.5	1312	2 US-08-592-126-148	Sequence 148, App
16	188	3.5	1312	2 US-08-687-080-51	Sequence 51, Appl
17	188	3.5	1312	4 US-09-168-595-148	Sequence 148, App
18	185	3.5	1972	4 US-08-875-435B-4	Sequence 4, Appli
19	183.5	3.4	1315	4 US-09-200-650E-5	Sequence 5, Appli
20	182	3.4	1166	4 US-09-200-650E-7	Sequence 7, Appli
21	180.5	3.4	840	3 US-08-974-549A-190	Sequence 190, App
22	180.5	3.4	840	4 US-09-402-181B-190	Sequence 190, App
23	180.5	3.4	840	4 US-09-721-456-190	Sequence 190, App
24	179	3.3	3248	1 US-08-353-700-1	Sequence 1, Appli
25	179	3.3	3248	5 PCT-US95-16216-1	Sequence 1, Appli
26	179	3.3	3418	2 US-08-603-753D-4	Sequence 4, Appli
27	179	3.3	3418	3 US-09-099-753-4	Sequence 4, Appli

28	179	3.3	3418	3 US-08-986-106-4	Sequence 4, Appli
29	177.5	3.3	1972	4 US-08-875-435B-3	Sequence 3, Appli
30	175	3.3	2285	4 US-09-308-375-2	Sequence 2, Appli
31	174.5	3.3	1027	3 US-08-446-137B-2	Sequence 2, Appli
32	174	3.3	3418	2 US-08-639-501-2	Sequence 2, Appli
33	174	3.3	3418	3 US-09-044-946-2	Sequence 2, Appli
34	174	3.3	3418	3 US-09-044-908-2	Sequence 2, Appli
35	173.5	3.2	1024	3 US-09-091-117-5	Sequence 5, Appli
36	173	3.2	2411	4 US-09-268-347-36	Sequence 36, Appl
37	171	3.2	3418	3 US-08-755-587-44	Sequence 44, Appl
38	169.5	3.2	1094	4 US-09-268-347-32	Sequence 32, Appl
39	169.5	3.2	1729	4 US-09-134-000C-5675	Sequence 5675, Ap
40	169	3.2	1664	1 US-09-599-652-2	Sequence 2, Appli
41	169	3.2	1664	2 US-08-642-846-2	Sequence 2, Appli
42	169	3.2	1664	4 US-09-264-604-2	Sequence 2, Appli
43	168.5	3.1	878	4 US-09-134-001C-4378	Sequence 4378, Ap
44	168.5	3.1	2343	3 US-09-324-867-2	Sequence 2, Appli
45	168.5	3.1	2954	4 US-09-150-867-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lyra Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064, 964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5080

Query Match	4.2%	Score 226.5;	DB 4;	Length 3696;
Best Local Similarity	19.6%	Pred. No. 1.3e-06;		
Matches 248;	Conservative 186;	Mismatches 459;	Indels 371;	Gaps 54;
QY	33 SVKQOTQSASEDDWFE-----EDNERKTNSKE--NSTYDETVSDLFSDGNSNSSSK 84			
DB	1242 NVKRRQEAKTINDIVQHKOSIQNDDATTEKEVANNLVNASQCNVISKIDNATNNQ 1301			
QY	85 TESVSDPKQVPKAKPEVTQASNSNDA-----SKVEVPKQDTASKKE-----TLRT 132			
DB	1302 IDGIYSDGRQISINAITPDTSIKNAKNIDIDIKAAKKIKIQINDATDEIEQANRKIEE 1361			
QY	133 STWEAKDFVTRGDT--LVGFSG--SGINKLSQ--TSHLVLPASHADGTQ--LTQVASFAFT 185			
DB	1362 AKIEAKDNIQNRSTRDQVNEAKNGINKIENITPATYVKSARQAQVNKANEQINHIQNT 1421			
QY	186 PD-----KKTALAEYTSRLGE-----NGKPSRLDIDQKEIIDEG-----EIFNAYQLTKLT 231			
DB	1422 PDATNEKQEAIRNVSAAELARVQAQINAHTTQGV--KTIKDAITSLSRINAQVVEKES 1479			
QY	232 IPNG--YKSIGQDAFVDN-----KNIAEVNLPESLETISDYAFAMSLKQVKL 277			
DB	1480 ARNAIEQKATQOTQFINNNDNATBEKEVANNLVIAATKQKSLDNIN-----SLSSND 1532			
QY	278 PDNLKVIIGELAFPDNIGGKLYLPRHLIKLAER-----AFKSNRIQTVFELGSKLKYI 330			
DB	1533 VENAKVAG-----INEIAN--VLPAATVSKAKKIDIDQKLAQOINQIQHTATTEEK-- 1583			

```
QY 331 GEASFDNNLRNMLPDGLEKIESEAFNGPGEDEHYNNQVYLKTRTGQNP HQ----- 383
Db 1584 -EALIQULANQKS-----NEARFALIQ-NEHSNNGVAQA KSNGLIHEIELVMPDAHK 1630
QY 384 -----ATENTYVNPDKSLWRATPD-----MDYTKMLEEDFTYQKNSVTGFSNKL 428
Db 1631 KSDAKQSIDNKY-NEQSNNTINTTPDATDEEKQALDKLK-IAKDAGYNK-----V 1678
QY 429 QKVRKNKLEIPKOHNGITITEIGDAFR---NVDFOSKTLKDYLEIK-----LPST 479
Db 1679 DQAQTMQOVSDAKTEAIDTITNIQANVAKKPSARVELDSK-----FEDLKRQINATPNA 1732
QY 480 IRKIGAFQSNMLKSFEAS-----EDLEEI-----KEGAFM-- 511
Db 1733 TEEKQDAIQRNGKRDEVKNLINODRRNEVEQHKNIGLQELTTHANPTRKSDALQEL 1792
QY 512 -----NNRIGTLDLKD-----LIKIGDAFH-----INH 537
Db 1793 QTKFISQTELLNNKDATNEEKDEAKRLLEISKNTITNINQAGTNNQVDNAKNGMNEI 1852
QY 538 YAIVLPEVSQEIGRSAPRONGALHLMFI-GNKVKTIGEMA----- 576
Db 1853 ATIIIPATTIKTDAKTAIDKKAEOQVTIINGNNDATDEEKAERKLVKAKIEAKSNTINS 1912
QY 577 -----FLSNKLESV-NLSEQKQKLTIEVQAFSDNALSEV---LPNLQITIREAF 623
Db 1913 DTEREVNGAKTNGLEKINNIOPTQGTAKNAQOEINDKAQEOLIQINNTPDATSEBKQEA 1972
QY 624 KR-----NHLKEVKGSSITLSQITFNAFDQNDGDKRFGKVVVTRHNSHM 668
Db 1973 NRVNAGLAQAIQINNAHSTOEVENESKTSIATIKSVQPNVIKPTAINSLTQEBANNQKT 2032
QY 669 LADGERFIIDPKLSSTWVDEKVLKITEGLDYSLRQTTQTQOFREMTAGKALLSKSNL 728
Db 2033 LIGNDGNATDEKEAAKQLVTOKLENOQIKIHSTQDNQVDNKAQAITAIKLINANAHK 2092
QY 729 ROGEKQKFLQEAQFPLGRVDLKAIAKAEKALVTKATKNGHLERSINKAV-----LA 782
Db 2093 RQ-----DAINILTNLAESKSDI--RANQDATTEEK--NTAIQSIDDTIA 2134
QY 783 YNNSAIKKANVKRLEKELDLTLTDLVEGKPLAQATWQGVYLLKTPLEPEVYTGANVYF 842
Db 2135 QARNNINGANTNALVDE---NLLEDGKQKQOR-----IVL 2165
QY 843 DKSGLIYALDMSPTIGEGQDAYGNPILNVEDNE-----GYHTLAVATLADYE 892
Db 2166 STQTKTQAKADIAQAIQOQRSTIDQONQATTEEKQEALERLNOETNGVNDRIQALANQN 2225
QY 893 GLYIKDILNSSLDKIKAIRQIPLAKYHRLGIPOAIRNAAAEADRLPKTPK----- 943
Db 2226 ---VTDEKNNTLETIRNVEPIVIVPKANEI---IRKKAEOFTLINQNDQATLEEKQIA 2279
QY 944 -GYLNEVPNYRKQKQMEKNLKPVDYKTPFN-----KALPNEKYVDGDRAAKG 988
Db 2280 LGKLEEVKNNEALNQVSAHSNNDVKIAENNGIAKISEVHPETIIKRNAKQEIHQDAQSOI 2339
QY 989 HNINA---ETNN-----SVA-----VTPIRSEQQLHKSQSDVN-----LPQTS 1023
Db 2340 DTINANNKSTNEKSAIDRVNAVAKIDAINNTINATTTQLVNDAKNSGNTSISQILPSTA 2399
QY 1024 SKNN 1027
Db 2400 VKTN 2403

RESULT 2
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 4.1%; Score 222; DB 4; Length 10182;
Best Local Similarity 19.7%; Pred. No. 1.2e-05;
Matches 236; Conservative 200; Mismatches 435; Indels 324; Gaps 58;

QY 29 ERESVQEQOTQASSEDWFEEDNERKTNVSKENSTVDETV---SDLFSDGNSNNSSKT 85
Db 8438 QREWINTNTNATTRREKAVAKNLDNAQA--LDKAMETLQGVAAHKNTILNDSKYLNEDSKY 8495
QY 86 ES-----VSDPKQVPKAKPEVTQEA SSSNDASKVEVPKODT-ASKK-----ETLETGTW 135
Db 8496 QQQYDRVIADAQ-----ILNQTNPTLEPYKVDIVKONVLANEKILFGAEKLSYDKS 8548
QY 136 EAKDVTYRGDTLVGFSKSGINKLSQTSHLVLP SHAADGTQLTQVASFATPDKKTAIAEY 195
Db 8549 NANDELKMNYYLNNAQKQSIKDM-----SHAALRFEVKQL----- 8584
QY 196 TSLRGNGKPSRLDIDQKEIIDEG---EIFNAVQLTKLTI PNGYKSIGOD----- 242
Db 8585 -----LQQA KILDEAMKSLDKTQVVTIDTLLPN-YTEASEDKKEKVDQTV 8629
QY 243 ---AFVDNKNIAEVLNLP-----ESLETISDYAFAHMSLKQVKLPDNLKVIGELAFEDN 292
Db 8630 SHAQAIIDKINGSNVSLDQVRQALEQLTQASENLDDGDQVVEAKVHAN-QTIDQLTHLNS 8688
QY 293 --QIGGKLYLPRHLIKLAERAFKSNRIQTVFELGSKLVIGEASFQDNNLRNMLPDGLE 350
Db 8689 LQQQTAHESV-KNATKLEELATVSNNAQALN-----KVMGKLEQFINHADSVENS DNYR 8741
QY 351 KIESEAFGT-NPGEDEHYNNQVNLRTTGQNP-----HOLATENTYVNPDKSLWRATP-DM 403
Db 8742 QADDKILAYDEALEH--GODIQKTNATQNETKQALQOLIYAFTSLNGFERLNHARPRAL 8799
QY 404 DYTQWLEEDFTYQKNSVTGFSNKGIOKVRRNKNLEIPKOH--NGITITEI-GDNAFRNV 459
Db 8800 EYIKSLEKINNAQKSALE-----DKYTQSHDL-LELEHIVNEGTLNDIMGELANAIIV 8851
QY 460 DFQSKTLRKVDLEIKLPSTIRKIGAFQSNNL-KSFEASEDIEEIKEGAFMN-NRIGT 517
Db 8852 NNYAPT-----KASINYINADNLKRDNFTQAINNARDALNKQGNLDENAI DT 8900
QY 518 LDKDKLIKIGDAAFHINHIYAI-----VLPEVSQEIGRSAP--RONGALHLMFIGNKVK 570
Db 8901 --FKDDIFKTRDALNGIERLTAAKSKAEKLIDSLKFINKAQFTHANDEIMNTNSIAQLSR 8958
QY 571 TIGEMAFLSNKLESVNLSEQKQKLTIEVQAFS-----DNALSEVVLPPNLQOT 617
Db 8959 IVNQAFDLNDAWKS--LRDELNNQAFPYQASSNYINSDEDLKQGFDAHLSNA-----RKV 9011
QY 618 IREAFKRNHLKEVKGSSSTLSQITFNAPQNDGDKRFGK--KVVVTRHNSHMLADGER 674
Db 9012 LAKENGKULDEKQIQG--LKQVIEDTKDALNGIQRLSKAKAKAIQYVQSISY-INDAQR 9067
QY 675 FTID-----PDKLSS-----TWVDEKVLK-----IIEGLDYSLRQTTQTO 711
Db 9068 HIAENNIHNSDLSLANLTSKASDLDNAMKDLRDTIESNSTSVPNVYINADKNLQIE 9127
QY 712 FREMTAGKALLSKS-----NLROGEKQKFLQEAQFPLGRVDLD 750
Db 9128 FDEALQASATSSKTSSENPATIEEVLGLSQAIYDTKNALNGE-QRLATEKSKDLKLIKGL 9186
```


QY	751	KAIKAKEALVTKKATNGHLLERS-----IN-KAVIAYNN	785
Dd	9187	KDLNKAQLDVTNKVNSANTTELSQTOSTLELNDKMKLLRDKLTVLNPVKASLYRN	9246
QY	786	S-----AIKKA-----NVKRLEKELDLLTDLVEGKPLAQATWQ	820
Dd	9247	ADYNLKRFNKALKEAKGYLNKNSGTWNINDIQHLLTQDNKDQLNGERRLKEHQKS	9306
QY	821	GYYLLKTPLPLEYYIGLN-----VYFDKSGKL-----YALDMSDTIGEQQDAY	866
Dd	9307	EVFIIK-----ELDILNNAQKAAIINOIRASKDIKIINOIVDNAIELNDAM-QGLKEHV	9359
QY	867	GN-----PILVNDEDEGXYHTLAVALADYEGLYIKDILNSLDKIKAIRQ---IP	914
Dd	9360	AQLTATTKDNIEYLNADEBHLQDYDAI-----NLANNVLDKENGNTNKDANI I	9407
QY	915	LAKYHRLGIFQAIRNPAAEADRLLPKTPKGYLINEVPNYRRKOMENLKPVDYKTPIFNK A	974
Dd	9408	IGMIQNMDARALLNG---IERLKDAQTKAH-NDIKDTLKRQLE----IEHANATSNSK	9459
QY	975	LPNEKVGDRAKG-HINTAETNSV-----AVTPIRSEQOLHKXSQSDVN	1018
Dd	9460	AQAKOMVEEARKALSNINDATSNDLVNOAKDEGSAIEHIHAD-ELPKAKLDAN	9513

RESULT 3

```

US-08-973-462-8
/ Sequence 8, Application US/08973462B
/ Patent No. 6191270
/ GENERAL INFORMATION:
/ APPLICANT: DRUIIHE, PIERRE
/ APPLICANT: DAUBERSIES, PIERRE
/ TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
/ FILE REFERENCE: 0660-0125-0 PCT
/ CURRENT APPLICATION NUMBER: US/08/973,462B
/ CURRENT FILING DATE: 1998-02-06
/ EARLIER APPLICATION NUMBER: PCT/FR96/00894
/ EARLIER FILING DATE: 1996-06-12
/ EARLIER APPLICATION NUMBER: FR 95/07007
/ EARLIER FILING DATE: 1995-06-13
/ NUMBER OF SEQ ID NOS: 29
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 8
/ LENGTH: 1786
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

```

Query Match	3.8%;	Score 201.5;	DB 3;	Length 1786;
Best Local Similarity	20.0%;	Pred. No. 2.7e-05;		
Matches 228;	Conservative 176;	Mismatches 433;	Indels 305;	Gaps 48;

[illegible]

```

Db      952  IPENLEENAVESNENVAE-NLEKLNETVENTVLD-----KVEETVEISGESLENNEMD 1003
QY      287  LAFFDNQIGKLYPRHLIKLAERAFKSN-RIQTVEEFLGSKLVIGEASFQDN--NLNV 343
Db      1004  KAFFSEIFDNVKGIOENLLTGMERSIETSIYIGEEKVDLNNVV--SSILDNIENMEKG 1061
QY      344  MLPDGLKIESEAFIGNPDDEHYNNQVVLRRITGONPHQLATENTYYNPPDKSLMRATPDM 403
Db      1062  IL-NKLENISSTEGVQETVTEH-----VEQNYYVDVD-----YFAM 1096
QY      404  DYTMLBEDFTYQKNSVTGFSNK--GLQXVRNKNLLEIPKOHNGITITEIGDNAFRYDF 461
Db      1097  -----KDQFLGILNEAGLGKEMFNLEDYFKSESVDITVEIKEYQVQK-EV 1142
QY      462  QSKTLRKYD-----LEEIKLPSTIRKIGA---FAQSNLKLKSFASED----- 501
Db      1143  EKETVSIIEEMEENIVDYLEEKEDLTDMKIDAVEESIEISSDSKETESIKDKEKDVSL 1202
QY      502  -LEEIKEGAFMNNRIGTLDKDLIKIGDAAFIHNIYAIVLPESVQEIGRSAFRQNGAL 560
Db      1203  VVEEVQDNMDMESVEKYLELKNMBEELMKDAVEINDITSKLI-EETQELNEV-----EA 1255
QY      561  HLMFIGNKVKTIGEMAFLSNKLBSVNLSEQKQKLTIEVQAFSDNALSEVVLPPNLQITRE 620
Db      1256  DLIKOMEKTEL-EKALSEDSKEIIDAQDTELEKVIIEEHIDITTLDEVV--ELKQVEE 1311
QY      621  EAFKR-----NHLKEVKGSSTL-SQI--TFNAFDONDGDKRFGKVVVRTH----- 663
Db      1312  DKIEKVSDDLKDLBEDILKEVKEIKELESEILDEYKELTITETDILBEKKEIEKDHPEKFE 1371

```

2

```

Db      1372  EEAEIKDLEADIL--KEVSSLEVEEKKLEEVHLEKEVEHIISGDAHIKGLLEEDDLEE 142929
QY      699  -----LDYSTLRQTTOQOFREMTTAGKALLSKSNLRQGEKQELQ 738
Db      1430  VDDLKGSILDMCLKGMELGDMKESLEDVT-TKLGERVESLKDVLSALGMDEEQMKTRK 14888
QY      739  EAQFELGRVDLDKAIKAE-----KALVTYKATK-----NGHLERSINKAV 780
Db      1489  KAQ---RPKLEEVLLKEEVKEBPKKKITKKKVRFDIKDKPKDEIVEVEMKDEIDIEDV 15444
QY      781  LAYNSAIKKANVKRLEKELDLLTDLVEGKGPILAQATM----VQGVYLLKTPL--PLPE 833
Db      1545  EEDIEEDIEEDKVEDIDEDID--EDIGEDKDEVIDLIVQEKRIEKVKAKKKLEKVEE 16022
QY      834  YYIGLNVYFDKSGKLIYALDMSDTIGEGOKDAYGNPILNVDEDNEGYHT----- 882
Db      1603  GVSGLKKHVDVVMKYQKIDKDEVDKEVSKALESKNDVYTNVLKONODFFSKVKNFVKYKY 16622
QY      883  -----LAVATLADY-EGLYIKDI-----INSLDK-IKAIRQIPLAKYHRLG 922
Db      1663  FAAPFISAVAAFAFASYVGFFTSLFSSCVTIASSTYLLSKVDKTINKNKERPFSF---- 17188
QY      923  IFQAIRNAAAEADRLLKPTPKGYLNEVENPYRRKKQOMEKULKPVDYKTPIFNKALPNEKYVDG 982
Db      1719  VFDFIKNLKHYLQOMEKEFSKEKNNNVIEVTNKAEEKGNVQVTNKTEKTTKVDKNNKVPK 17788
QY      983  DR 984
Db      1779  KR 1780

```

RESULT 4

```

US-08-851-843A-8
; Sequence 8, Application US/08851843A
; Patent No. 6093809
;
; GENERAL INFORMATION:
;
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morfin, Gregg B.
;

```

APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. 6093809el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 872 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: protein
 US-08-851-843A-8

Query Match	3.7%;	Score 200;	DB 3;	Length 872;
Best Local Similarity	20.3%;	Pred. No. 1.2e-05;		
Matches 200;	Conservative 143;	Mismatches 350;	Indels 294;	Gaps 48;

OY		14	TTVSAVTYSQEVYLE-----REESVK-----QEOTQSASEDDWFEEDNERK-	55
Dp		17	TNUDFVLONLEVYKSQIEHYKTQQQIKBEDLKLKEKNQDQDGNSGNDDEENSNKO	76
OY		56	-----TNVSKENSTVDETVSDLFSPDGNSSNNSSKTESVSDPKOVPAKAEVTQEASN	108
Dp		77	QELLRRVNQIKQOVQLIKKVSGSKEKDNLN-----EDEN	111
OY		109	SSUNDASKVEVPKODTASKKETLETSTWEA--KDFVTRGDTLVGFSKG----	154
Dp		112	KKNGISEQOY-----KEEQLRITTEEOVKYONLVFNMDYOOLDINESGGHRRHRRET DY	164
OY		155	-INKLSQTSHLVLPSHAADGTOLTQVASFAFTPDKKTAIAEYTSRLGENGKPSRDLIDOK	213
Dp		165	DTEKMFEISH-----DOKNYVSIVA---NQKTSYCWWLKDYFNKKNYDHNV SIN	211
OY		214	EIIDEGELF--NAYOLTKLTIPNGYKSIGODA FVDNK-----NIAEV-	253
Dp		212	RLETFEAFFYA FDDFSQTI KLT- NNSYTQTVANIVAFDN NLCLALLRFLSLERENILNI R	270
OY		254	-----NLPSELETISDYAFAHMSLKQVKLPDNLKVI GELAFFDNQIGK-----	297

Db	271	SSYTRNÖYNFEKIGELLEITIFAVVFESHRLÖGIMHÖVPCEAFQYLVNSSSÖISVKDSÖLÖ	330
QY	298	LY-----LPRHLIKLAERAFKSNRIQVEFLGSKLVIGEAS	334
Db	331	VYSEFSTDKLVDITNKQDYEFKPLÖEFPRLTHVSÖQALPVSAITNAVENLNLKKVKHA-	388
QY	335	FÖDNNLRNVMPLDGLKIESEAFGTGNPGDEHYNNÖVILKTRIGÖNPÖLATENTYVNPDK	394
Db	389	-----NLNLVSIP-----TÖFNEDFYFVNLÖHLEKLEFGHEPNILTKÖKL-----E	428
QY	395	SLMRATPDMDYTKWLEEDF-TY--ÖKNSVTGFSNKG--LÖKYBRNKNLEIPKÖHNGITIT	449
Db	429	NLLLSIKÖSKNLKFRLNRNFYTYVAÖETSRKÖILKÖATTIKNLKNKNQÖETPETKDETPS	488
QY	450	EIGDNAFRNVDFOŠKTLRKXYDLBEIKLPST---IRKIGAFAFÖSNNLKSFEAS--EDLEE	504
Db	489	E-STSGMKFEDHLSLTELDEF-SVNLQATÖEIDYSLHKLILRSTNULKFEKLSYKYEEMK	546
QY	505	IKEGAF--MNNRIGTL-DLKDKLIKIGDAFHINHIYAIVLPESV-----QEI--G	550
Db	547	SKMDTFIDLKNIYETLNNLKRCSVNIŠNP--HGNISYELTNKDSGFYKFKLTLLNÖELÖHA	604
QY	551	RSAPFRÖNGALHLMFIGNKVKITIGEMAFLSNKLBSV-----NLSEÖKÖLTIHVQAFS	602
Db	605	KYTEKÖN---EFÖFNNVKS--AKIESSSLESLEDIDSLCKSIASCNÖLÖNV-----	650
QY	603	DNALSEVVLPPNLÖTIRBEAFKRNHL--KEVKGSSSTLSÖITENA-FDÖNDGDGRFGKKV	658
Db	651	-NIASLLYPNNIÖ--KNPFKNPULLFFKÖFÖLKNLENVSINCILDÖ-----	695
QY	659	VVRTHNŠHMLADGERFIIDPDKLSTWVDLEKVLKIEGLDYŠTLRÖTQ-----	709
Db	696	-----HILNISSEFLEKNKKIKAFILKRYVLLÖYY--LDYTKLFKTLÖQÖLPENÖVY	745
QY	710	--TÖPREMTTAGKALLSKSNLROŠEKÖK-----FLÖEAÖFGLGRVIDDKAIA---KA	756
Db	746	INÖÖLEELTVSE---VHKÖVWENHÖKAFYEPPLCEFIKESSÖTLÖLIDPÖQNTVSDSI	801
QY	757	EKALVTKKAITKNGHLLERSINKAVLAVNNSAIKKANVKRLLEKELDLTDLVEGKPLAÖA	816
Db	802	KKILIESISEŠKYHHYLRNPŠÖ-----SSSLIKŠEN-----EIÖELLKACDEKGVLVKA	851
QY	817	TWVÖGVYLLKTPPLPPEYYIGLVNYFD	843
Db	852	-----YKFPPLCLP-----TGTYYD	866

RESULT 5

US-08-851-843A-54
; Sequence 54, Application US/08851843A
; Patent No. 6093809

1 GENERAL INFORMATION:
2 APPLICANT: Cech, Thomas R.
3 APPLICANT: Lingner, Joachim
4 APPLICANT: Nakamura, Toru
5 APPLICANT: Chapman, Karen B.
6 APPLICANT: Morin, Gregg B.
7 APPLICANT: Hatley, Calvin
8 APPLICANT: Andrews, William H.
9 TITLE OF INVENTION: No. 6093809e1 Telomerase
10 NUMBER OF SEQUENCES: 225
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Townsend and Townsend and Crew LLP
13 STREET: Two Embarcadero Center, 8th Floor
14 CITY: San Francisco
15 STATE: California
16 COUNTRY: United States of America
17 ZIP: 94111
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-851-843A-54

Query Match 3.7%; Score 200; DB 3; Length 872;
Best Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY 14 TTTSVVTYSQEVYGLF-----REESVK-----QEQTOSASEDDWFEEDNERK- 55
DB 17 TNLDFVLONLEVYKSQLIHYKTQQQOIKEEDKLKRNQDQDGNSSGNDDEENNSNKQ 76
QY 56 -----TNVSKENSTVDETVSDLFSDGNSNNSSTKTESVSDPKQVPKAKPEVTQEASN 108
DB 77 QELLRRVNOIKQOVOLIKVGSKEVDLNLN-----EDEN 111
QY 109 SSNDASKVEVPKQDTASKKETLETSTWEA---KDFYTRGDTLVGFSKSG----- 154
DB 112 KKNGLSEQOV-----KEQLRTITEEQVKYQNLVENMDYQLDINESGHRHRETDY 164
QY 155 -INKLSQTSHLVLPASHADGTLQVASFATPPDKKTAIAEYTSRLGENGKPSRLDIDQK 213
DB 165 DTEKMFESH-----DQKNVYSIYA---NQKTSYCWMLKDYFNKNKNYDHLNVSIN 211
QY 214 EIIDEGEIF---NAYOLTKLTPNGYKSGIGDAFVNDK-----NIAEV- 253
DB 212 RLETEAEFYAFDDFSQTIKLT-NSSYQTVNIDVFNNDNLCLALLRFLSLERFNILNIR 270
QY 254 -----NLPESELETSDYAFAMSLKQVKLPDLKVIYIGELAFDNOIGK----- 297
DB 271 SSYTRNOYNFEKIGELLETIPAVVFSHRHLQGIHLQVPCAFQYLVNSSSQISVDSOLO 330
QY 298 LY-----LPRHLIKLAERAFKSNRIQTVEFLGSKLKVIYGEAS 334
DB 331 VYGFSTDLKLVDTNKVQDYFKPLQEPFR-LTHVSQQAIPYSATNAVENLNLTKKYKHA- 388
QY 335 PQDNNLRNVMLPDGLEKIESAFTGNPGEHYNQVLRTRTGQNPQOLATENTYVNDK 394
DB 389 ---NLNLVSIPI-----TQFNFDYFVNLQHLKLEFGLPEPILTKQKL-----E 428
QY 395 SLWRATPMDYTKMLEEDF-TV--QKNSVTGFSNKG--LQKVRNKNLEIPKQHNIGITIT 449
DB 429 NLLLSIKQSKNLKFLRLNFYTYAAQETSRSKQILKQATTIKNLKNNKNQEEPTETKDEPFS 488

QY 450 EIGNAFRNVDFQSKTLRKXYDLEIKLPST---IRKIGAFAFQSNLKSFEAS--EDLEE 504
DB 489 E-STSGMKFEDHSELTELED-F-SVNLQATQEIYDSLHKLILRSTNMLKFKLSYKEMEX 546
QY 505 IKEGAF--MNNRIGTL-DLKDKLIKIGDAAFHINHIAIVLPESV-----QEI--G 550
DB 547 SKMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSIFYKFKLTINQELQHA 604
QY 551 RSAFRONGALHMFIGNKVKITIGEMAFLSNKLSEV-----NLSEQOLKTIIEVQAFS 602
DB 605 KYTFKQN-----EFQFNNVKS--AKIESSSLESLEDIDSLCKSIASCNQLQNV----- 650
QY 603 DNALSEVVLPPNLQITIREBAFKRNHL---KEVGSSTLSQITFNA-EDQNDGDKRFGKV 658
DB 651 -NIIASLLYPNNIQ--KNPFKNPILFEKQFEQLKNLENVSNICLIDQ----- 695
QY 659 VVRTHNNSHMLADGERFIIDPKLSSTWVDLEKVLKIIEGLDYSTLRQTO----- 709
DB 696 -----HILNSISEFLEKNKKIKAFILKRYLLQYY--LDYTKLFKTLQQLPELNOVY 745
QY 710 --TQFRMTAGKALLSKSNLRQGEKQK-----FLQEAQFLLGRVLDKAIA---KA 756
DB 746 INQLEELTVSE-----VHKQWENHKKQKAFYEPLCEFIKESQTLQLDIDQNTVSDSI 801
QY 757 EKALVTKKATKNGHLLERSINKAVLAVNNSAIKANKVRLKEKELDLJDLVEGKPLAQ 816
DB 802 KKILSISESKYHYRLNPSQ-----SSSLIKSEN-----EIQELKLACDEKGLVKA 851
QY 817 TMVQGVYLLKTPPLPEYYIIGLVNYPD 843
DB 852 -----YKFPPLCLP-----TGTYYD 866

RESULT 6
US-08-974-549A-221
Sequence 221, Application US/08974549A
Patent No. 6166178
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:


```

; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph Ted
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002610US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 221:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-974-549A-221

```

Query Match	3.7%;	Score 200;	DB 3;	Length 872;
Best Local Similarity	20.3%;	Pred. No. 1.2e-05;		
Matches 200;	Conservative 143;	Mismatches 350;	Indels 294;	Gaps 48

QY	14	TTVSVVTVYSQOEVYGLE-----REESVK-----QEOQTOSASEDDMFEEEDNERK-	55
		: : : : : : : : : :	
Dp	17	TNIDFVLONLÉVYKKSQIEHYKTQOQOIKEEDIKLILFKFNQDQDGNSGNDDDEENNNSNKQ	76
QY	56	-----TNVSKENSTVDETVSDLFSDGNSNNSSSKTESVSVDPKQVPEKAPPEVTQOEASN	108
		: : : : : : : :	
Dp	77	QELLRRVNQIKQOVOLIKKVGSKVEKDNLN-----EDEN	111
QY	109	SSNDASKVEVPKQDTSKKEETLETSTWEA---KDFVTRGDTLVGFSKSG-----	154
		: : : : : : : : :	
Dp	112	KKNGISEQOV-----KEBQLRTITEEQVKYONLVFNMDYQLDINESGHRHRRET DY	164
QY	155	-INKLSQTSHLVLP SHAADGTQLTQVASFAPTPDKKTALAEYTSRLGENGKPSRLIDQK	213
		: : : : : : : : :	
Dp	165	DTEKWEFESH-----DQKNVYSIYA---NQKTSYCWMLKDYFNKNNDHLNVSIN	211
QY	214	EIIDEGEIF--NAYQLTKLTIPNGYKSIGODAFVDNK-----NIAEV-	253
		: : : : : : : : :	
Dp	212	RLTEAEAFYAFDDFSQTIKLT--NNSYQTVNIDVNFNNLCLIALLRFLSLERFININIR	270
QY	254	-----NLPESETISDYAFAHMSLKQVKLPDLNKVIGELAFFDNQIGK----	297
		: : : : : : : :	
Dp	271	SSYTRNQYNEFEKIGELLETFI FAVVFESHRLQGIHLQVPCFAFYLVNSSSQISVXDSQLQ	330
QY	298	LY-----LPRHLIKLAERAFKSNRIQTVEFLGSKLVIGEAS	334
		: : : : : : : : :	
Dp	331	VYSFSTDLKLVDTNKVQDYFKFLQOEPR-LTHVSQOALPVSATNAVENTNLKKVKGHA-	388
QY	335	FQDNNLRNVMLPDGLKIESEAFGTGNPGDEHYNNQVLRTRTGQCNPHOLATENTYVNEBK	394
		: : : : : : : : :	
Dp	389	----NLNLVSI-----TQFNFDFFYFVNLOHLKLEFGLEPNILTKQKL-----E	428
QY	395	SLWRATPMDYTKWLEEDF-TY--QKNSVTGFSNKG--LQKVRANKNLLEIPKOHNGITTT	449
		: : : : : : : : :	

```

Db      429  NLLLSIKOSKNLKFRLRLNFYTYVAQETS RKQILKQATTIKNLKNNNQOETPETKDETPS 488
QY      450  EIGDNAFRNVDFQSKTLRKXDL EEIKLPST---IRKIGAFAPQSNNUKSFEAS--EDLEE 504
Db      489  E-STSGMKFFDHLSELTELEDF-SVNLQATQEIYDSIHLKLLIRSTNLKKEFLSYKYEMEK 546
QY      505  IKEGAF--MNNRIGTL-DLKDKLIKIGDAAFHINHIAIVLPESV-----QEI--G 550
Db      547  SKMDTFIDLKNITYETLNNLKRCSVNISNP--HGNISYELTNKSDSTFYKFKTLNQELQHA 604
QY      551  RSAFRONGALHLMFIGNKVYKTIGEMAFLSNKLSEV-----NLSEQOLKTEIVQAFS 602
Db      605  KYTFKQN----EFQFNNVKS--AKIESSSLESLEDIDSLCKSIASCNLQNV----- 650
QY      603  DNALSEVLPENLQITIREAEFKENHL---KEVKGSSSTLSQITFNA-FDQNDGDKRFCKKV 658
Db      651  -NITASLLYPNNIQ--KNPFKNPLLEFKQFQOLKNLENVSINCLIDQ----- 695
QY      659  VYRTHNNSHMLADGERFIIDPKLISTMWDLKVLKIEGLDYSTRQTO----- 709
Db      696  -----HILNSISEFLEKNKKIKAFILKRYVLLQYY--LDYTKLFKTLQQLBELNQVY 745
QY      710  --TOFREMTTAGKALLSKSNLRQGEKQ-----FLQEAQFLLGRVLDPKAIA---KA 756
Db      746  INQQLBELTVSE---VHKQWENHHQKAFYEPLCEFIKESSQTLQLIDFDQNTVSDDSI 801
QY      757  EKALVTKKATKNGHLLERSINKAVLAYNNSAIKKANVKRLKELDLTLTDLVEGKPLAQ 816
Db      802  KKLIESISESKYHHYLRLPMSQ-----SSLLIKSEN-----EIQELLKACDEKGVLVKA 851
QY      817  TMOQGVLLKTPLPPEYYIGLVNYFD 843.
Db      852  -----YKFPPLCLP-----TGTYVD 866

```

```

1  RESULT 7
2  US-08-854-050-8
3  ; Sequence 8, Application US/08854050
4  ; Patent No. 6261836
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Cech, Thomas R.
7  ; APPLICANT: Lingner, Joachim
8  ; APPLICANT: Nakamura, Toru
9  ; APPLICANT: Chapman, Karen B.
10 ; APPLICANT: Morin, Gregg B.
11 ; APPLICANT: Harley, Calvin
12 ; APPLICANT: Andrews, William H.
13 ; TITLE OF INVENTION: No. 6261836el Telomerase
14 ; NUMBER OF SEQUENCES: 225
15 ; CORRESPONDENCE ADDRESS:
16 ; ADDRESSEE: Townsend and Townsend and Crew LLP
17 ; STREET: Two Embarcadero Center, 8th Floor
18 ; CITY: San Francisco
19 ; STATE: California
20 ; COUNTRY: United States of America
21 ; ZIP: 94111
22 ; COMPUTER READABLE FORM:
23 ; MEDIUM TYPE: Floppy disk
24 ; COMPUTER: IBM PC compatible
25 ; OPERATING SYSTEM: PC-DOS/MS-DOS
26 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
27 ; CURRENT APPLICATION DATA:
28 ; APPLICATION NUMBER: US/08/854,050
29 ; FILING DATE: 09-MAY-1997
30 ; CLASSIFICATION: 536
31 ; APPLICATION DATA:
32 ; APPLICATION NUMBER: US 08/851,843
33 ; FILING DATE: 06-MAY-1997
34 ; CLASSIFICATION: 536
35 ; PRIOR APPLICATION DATA:
36 ; APPLICATION NUMBER: US 08/846,017
37 ; FILING DATE: 25-APR-1997
38 ; CLASSIFICATION: 536

```

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER:  US 08/844,419
3      FILING DATE:  18-APR-1997
4      CLASSIFICATION:  536
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER:  US 08/724,643
7      FILING DATE:  01-OCT-1996
8      CLASSIFICATION:  536
9      ATTORNEY/AGENT INFORMATION:
10     NAME:  Apple, Randolph T.
11     REGISTRATION NUMBER:  36,429
12     REFERENCE/DOCKET NUMBER:  015389-002930US
13     TELECOMMUNICATION INFORMATION:
14     TELEPHONE:  (415) 576-0200
15     TELEFAX:  (415) 576-0300
16     INFORMATION FOR SEQ ID NO:  8:
17     SEQUENCE CHARACTERISTICS:
18     LENGTH:  872 amino acids
19     TYPE:  amino acid
20     STRANDEDNESS:  not relevant
21     TOPOLOGY:  not relevant
22     MOLECULE TYPE:  protein
23
24     US-08-854-050-8

```

Query Match	3.7%;	Score 200;	DB 3;	Length 872;
Best Local Similarity	20.3%;	Pred. No. 1.2e-05;		
Matches 200;	Conservative 143;	Mismatches 350;	Indels 294;	Gaps 48;

QY	14	TVTVSVTVTYSQE	VYGLE	-----	REESVK	-----	QEOTQASASEDDW	EE	DN	ERK	55																																																
Db	17	TNLD	FVLQNL	EVYKSQ	IEHYKTQQQ	QIK	IEDLKL	IFKNQD	QD	GN	SGND	DE	NN	SNKQ	76																																												
QY	56	-----	TN	SKEN	ST	VD	ET	VS	DL	PS	DG	SN	SS	KTES	V	SD	PKQV	PKA	KE	VT	Q	E	ASN	108																																			
Db	77	QEL	LR	VN	QIKQ	OVOL	IK	KG	V	KE	DL	N	-----	EDEN	111																																												
QY	109	SN	DA	SK	VE	VP	KQ	D	T	A	S	K	E	T	E	T	S	T	W	E	A	-----	KD	F	T	R	G	D	T	L	V	G	F	S	K	G	-----	154																					
Db	112	KN	GS	EQ	CV	-----	KE	Q	L	R	T	I	T	E	E	O	V	K	Q	N	L	V	E	N	M	D	Y	Q	L	D	L	N	E	S	G	H	R	H	R	E	T	D	Y	164															
QY	155	-	IN	L	S	Q	T	S	H	L	V	L	P	S	H	A	D	G	T	Q	L	T	O	V	A	S	F	A	P	T	P	D	K	K	T	A	I	A	E	Y	S	R	L	G	E	N	G	K	P	S	R	L	D	I	D	Q	213		
Db	165	D	T	E	K	M	E	I	S	H	-----	DQ	K	N	Y	S	I	T	A	-----	NQ	K	T	S	C	W	M	L	K	D	Y	F	N	K	N	N	Y	D	L	N	S	I	N	211															
QY	214	E	I	D	E	G	E	I	F	-----	N	A	Y	Q	L	T	K	L	T	I	P	N	G	K	S	I	G	O	D	A	F	V	D	N	K	-----	N	I	A	E	V	-----	253																
Db	212	R	L	E	T	E	A	E	F	Y	A	P	D	D	S	Q	I	K	L	T	-----	N	N	S	Y	Q	I	V	A	N	I	D	V	N	F	D	N	N	L	C	I	L	A	L	R	F	L	S	L	E	R	E	N	I	L	N	I	R	270
QY	254	-----	N	L	P	E	S	L	E	T	I	S	D	Y	A	F	A	H	S	L	K	O	V	K	L	P	D	N	L	K	V	I	G	E	L	A	F	E	D	N	O	I	G	K	-----	297													
Db	271	S	S	T	R	N	Q	N	Y	N	F	E	K	I	G	E	L	L	E	T	I	F	A	V	F	S	H	R	H	L	Q	I	H	L	O	V	P	C	E	A	F	O	Y	L	V	N	S	S	Q	I	S	V	K	D	S	O	L	Q	330
QY	298	L	Y	-----	-----	L	P	R	H	L	I	K	L	A	E	R	A	F	K	S	N	R	I	Q	T	V	E	F	L	G	S	K	I	V	I	G	E	A	S	334																			
Db	331	V	S	P	S	T	D	L	K	L	V	D	T	N	K	V	O	D	Y	E	K	F	L	Q	E	F	P	R	-----	L	H	V	S	Q	A	I	P	S	A	T	N	A	V	E	N	L	V	L	K	K	Y	K	H	A	-----	388			
QY	335	F	O	D	N	L	R	N	V	M	L	P	D	G	L	E	K	I	E	S	E	A	F	T	G	N	P	G	D	E	H	Y	N	N	Q	V	L	R	T	R	I	G	N	P	H	O	L	A	T	E	N	T	Y	V	N	P	D	K	394
Db	389	-----	N	L	N	L	V	S	I	P	-----	T	Q	F	N	D	E	F	V	N	L	Q	H	L	K	E	F	G	L	E	P	N	I	L	T	Q	K	L	-----	E	428																		
QY	395	S	E	M	R	A	T	E	P	D	M	D	Y	T	K	W	L	E	D	F	-----	T	Y	-----	Q	K	N	S	T	A	G	S	N	K	G	-----	L	Q	K	V	R	R	N	K	N	L	E	I	P	K	O	N	G	I	T	I	T	449	
Db	429	N	L	L	S	I	K	O	S	K																																																	

Db	605	KYTFKÖN-----EFÖFNANYKS---AKIESSSLESLEDIDSLCKSIASCNQLÖNV-----	650
QY	603	DNALSEVLPNLPÖTIREAEKRNHL---KEVKGSSSTLSÖITFNA-EDÖNDGDKRFEGKV	658
Db	651	-NIASLTYPNNIÖ---XNPNKNPMLLEFKÖFÖLKNLENVSINCILDÖ-----	695
QY	659	VVRTHNNSHMLADGERFIDDPDKLSTMWDLKVLKIEGLDYSTLRÖTTÖ-----	709
Db	696	-----HILNISEFLEKNKKIKAFILKRYLLÖY--LDYTKLFKTLÖÖLPENÖVY	745
QY	710	--TÖFREMTAGKALLSKSNLRÖGEKÖK-----FLÖEAÖFPLGRVIDLKAIA---KA	756
Db	746	INÖÖLEELTVSE---VHKÖWENHHKÖKAFYEPLCEFIKESSÖTLÖLIDEDÖNTVSDSI	801
QY	757	EKALVTKKATKNGHLLERSINKAVLAYNNSAIKKANVKRLKEKELDILTLVEGKPLAÖA	816
Db	802	KKILESISESKYHHYLRNLPSÖ-----SSLLIKSEN---EÖIÖELLKACDEKGVLVKA	851
QY	817	TMVÖGVYLLKTPLPPLPEYYITGLNVYFD	843
Db	852	-----YKFPPLCLP-----TGTYVD	866

RESULT 8
US-08-854-050-54
Sequence 54, Application US/08854050
Patent No. 6261836
GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6261836el Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429


```

Db      165 DTEKWEISH-----DQKNVSIYA---NOKTSYCWMLKDYFNKNVYDHLNVSIN 211
QY      214 EIIDEGEIF---NAYQTLKLTIPNGYKSIGQDAFVDNK-----NIAEV- 253
Db      212 RLETEAEFYAFDDFSQTIKLT--NNSYQTVNIDVNFDMNLCTIALRLFLSLERFNILNIR 270
QY      254 -----NLPESLETISDYAFAMSLKQVKLPDNLKVIIGELAFPDNQIGK----- 297
Db      271 SSYTRNQYNFEKIELETTIFAVVESHRLQGHQVCEAFQYLVNSSQISVDSQLO 330
QY      298 LY-----LPRHLIKLAERAFKSNRIQTVFELGSKLVIGEAS 334
Db      331 VYSFSTDLKLVDTNKVQDYFKFLQEFPR-LTHVSQQAIPVSATNAVENVLLKKVKAH- 388
QY      335 FQDNMLRNVMMLPDGLEKIESEAFNGPDEHYNNQVLRTRTGQNPHQLATENTYVNPDK 394
Db      389 -----NLNLVSI-----TQNFDFYFVNLQHLKLEFLEPNILTKQKL-----E 428
QY      395 SLMRATPMDYTKWLEEDF-TY--QKNSVTGFSNKG--LOKVRANKNLEIPKQNGITIT 449
Db      429 NULLSIKQSKNLKFLRLNYYTYVAQETSRLQKQATITIKLNKNKNOEETPETKDETPS 488
QY      450 EIGDNAFRNVDFQSKTKRYDLEIKLPST--IRKIGAFQFQSNMLKSFAS--EDLEE 504
Db      489 E-STGSKMFPDHLSELTELEDF-SVNLQATQEIYSLHKLILIRSTNLKFKLSYKYEEMK 546
QY      505 IKEGAF--MNNRIGTL-DIKDLIKIGDAFHNHYAIVLPESV-----OEI--G 550
Db      547 SKMDTFIDKNIYETLNNLRKCSVNISNP--HGNISYELTNKSDTFYKFKLTINQLOHA 604
QY      551 RSAFRONGALHLMFIGNKYTIGEMAFLSNKLSEV-----NLSEOKQKLTIEVQAFS 602
Db      605 KYTFKQN-----EFQFNNVKS---AKIESSLESLEDIDSLCKSIASCNNLQNV----- 650
QY      603 DNALSEVLPNLTQITREAFKRNHL---KEVKGSSSTLSQITENA-FDQNDGDKRFGKV 658
Db      651 NIIASLTPNNIQ--KNPENKNLFFKQFEQKLENVNSINCILDQ----- 695
QY      659 VVRTHNNSHMLADGERFIIDPKLSTWVDEKVLKIEGLDYSTRQTO----- 709
Db      696 -----HILNSISEFLEKKNKIKAFILKRYLLOY--LDYTKLFKTLQQLPELNQY 745
QY      710 --TQREMTAGKALLSKSNLRQGEKQK-----FLQEAQFPLGRVLDKATA--KA 756
Db      746 INQOLEELTVSE---VHKQWENHKQKAFYEPLCEFIKESSTQLQIDFDQNTVSDSI 801
QY      757 EKALVTKKATKNGHLERSINKAVLAYNNSAIKKNVRLKEKELDLTLVSEGKPLAQ 816
Db      802 KKLIESISESKYHHLRLNPSQ-----SSSLIKSEN-----EIQELKACDEKGLVKA 851
QY      817 TWVQGVYLLKPLPLEYIIGLVNVPD 843
Db      852 -----YKFPPLCLP-----TGTYYD 866

```

RESULT 11
US-09-402-181B-221

; Sequence 221, Application US/09402181B
; Patent No. 6610839

; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.

; Lingner, Joachim
; Nakamura, Toru

; Chapman, Karen B.
; Morin, Gregg B.

; Harley, Calvin B.
; Andrews, William H.

; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 633
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor

```

; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/402,181B
; FILING DATE: 29-Sep-1997
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ausenhus, Scott L.
; REGISTRATION NUMBER: 42,271
; REFERENCE/DOCKET NUMBER: 015389-002620US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 221:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 221:
US-09-402-181B-221

Query Match      3.7%; Score 200; DB 4; Length 872;
Best local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY      14 TTTSVTVYSQEVYGLF-----REESVK-----QEQTQASGDDWFEEDNERK- 55
Db      17 TNLDFVLQNLLEVYKQIEHYKTOQQQIKEDLKLKPKNQDQDNGSNDDEENNSNKQ 76
QY      56 -----TNVSKNSTVDETVSDLFSDGNSNNSKTESVSDPKQVPAKPEVTQEAQN 108
Db      77 QELRRVNGIKQOVOLIKKVGSKVEKDLNIN-----EDEN 111
QY      109 SSNDASKEVPEKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG----- 154
Db      112 KKNGLSEQQV-----KEQRLRTITEQVYQYNLVFNMDDYQDLNBSGGHRRHRETQY 164
QY      155 -INKLSQTSHLVLPASHADGTQLTVASFAFTPDKKTALIAEYTSRLGENGKPSRLDIDQK 213
Db      165 DTEKWEISH-----DQKNVSIYA---NOKTSYCWMLKDYFNKNVYDHLNVSIN 211
QY      214 EIIDEGEIF---NAYQTLKLTIPNGYKSIGQDAFVDNK-----NIAEV- 253
Db      212 RLETEAEFYAFDDFSQTIKLT--NNSYQTVNIDVNFDMNLCTIALRLFLSLERFNILNIR 270
QY      254 -----NLPESLETISDYAFAMSLKQVKLPDNLKVIIGELAFPDNQIGK----- 297

```

```

Db      271 SSYTRNQYNFEKIGELLETFEAVVFSHRHLOGIHLOVPCFAFOYLWNSSSQISVKDSQLO 330
QY      298 LY-----LPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGEAS 334
Db      331 VYGFSTDLKLVDTNKVQDYFKPLQEFPR-LTHVSQQAIPVSAFMVAVENLVLTKVYKA- 388
QY      335 FQDNNLRVMLPDGLEKIESAFTGNPGDEHYNNQVLRFTGQNPQOLATENTYVNDK 394
Db      389 ----NLNLVSI-----TQNFDFEYVNLQHLKEFGLEPNILTKQKL-----E 428
QY      395 SLWRATPMDYTKWLEEDF-TY--QKNSVTGFSNKG--LQKVRNNKMLEIPKQHNGITIT 449
Db      429 NLISIKQSKNLKFLRLNFYTYVAQETSRKQILKQATTIKNLKNNKNOBETPETKDETPS 488
QY      450 EIGDNAFRNVDFOSKTLRKVDIEIKLPST--IRKIGAFQSNLKSFEAS--EDLEE 504
Db      489 E-STSGMKFEDHSELTELEDF-SVNLQATQEIYDSLHKLILRSTNLKKFKLSYKEMEK 546
QY      505 IKEGAF--MNNRIGTL-DLKDKLIKIGDAFHINHYAVLPESV-----QET--G 550
Db      547 SKMDTFIDKNIYETLNLRKCSVNISNP--HGNISYELTNKDSFTYKFKLTNLQLOHA 604
QY      551 RSAFRONGALHLMFIGNKVTIGEMAFLSNKEV-----NLSEQKQKLTIEVOAFS 602
Db      605 KYTRKQK----EFQFNNVKS--AKIESSSLESLEDIDSLCKSIASCKMLQNV----- 650
QY      603 DNALSEVLPNLOTIREAFKRNHL--KEVKGSSTLSQITENA-FDQNDGDKRFGKV 658
Db      651 -NIIASLTPNNIQ--KNPKNPNLLFKQFEQLKMLNVSINCILDQ----- 695
QY      659 VVRTHNNSMLADGERFIDDPKLSSTWDLKVLKITEGLDYSTLRQTTQ----- 709
Db      696 -----HINSISEFLEKNKKIKAFILKRYLLQYY--LDYTKLFTLQQLPELNQVY 745
QY      710 --TOREMTTAGKALSKSNLRQGEKQK-----FLQEAQFELGRVDLDKATA--KA 756
Db      746 INQOLEELTVSE---VHKQVWENHKQKAFYEPLCEFTIKESSQTLQIDFDQNTVSDSI 801
QY      757 EKALVTKKATKNGHLERSINKAVLAYNNSAIKKANVRLEKELDLITDLVEGKPLAQ 816
Db      802 KKLIESISESKYHHYLRNLSQ-----SSSLIKSEN-----EIQELLKACDEKGVLYKA 851
QY      817 TMVQGVYLLKTPLPPEYVYIGLVYFD 843
Db      852 -----YKFPPLCLP-----TGTYVD 866

```

RESULT 12
US-09-721-456-221

Sequence 221, Application US/09721456
Patent No. 6617110

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

Lingner, Joachim

Nakamura, Toru

Chapman, Karen B.

Morin, Gregg B.

Harley, Calvin B.

Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727

CORRESPONDENCE ADDRESS:

ADDRESSER: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-Nov-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-AUG-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-OCT-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 872 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 221:
US-09-721-456-221

Query Match      3.7%; Score 200; DB 4; Length 872;
Best Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 200; Conservative 143; Mismatches 350; Indels 294; Gaps 48;

QY      14 TTVSVTYTSQEVYGLF-----REESYK-----QEQTQASASEDDWFEEDNERK- 55
Db      17 TNLDFVLQNLLEVYKQIEHYKTQQQIKKEEDLKILKFNQDQDGNSGNDDDEHNSNKQ 76
QY      56 -----TNVSKENSTVDETVSDLFSDGNSNNSSSKTESVSDPKQVPKAKPEVTQEASN 108
Db      77 QELLRVNLQKQVQLIKKVGSKVEKDILNLN-----EDEN 111
QY      109 SSNDASKVEVPKQDTASKKETLETSTWEA---KDFVTRGDTLVGFSKSG----- 154
Db      112 KKNGLSEQV-----KEQRLRTITEEQVKQNLVENMDYQLDLNESGGHRRHRETDTY 164
QY      155 -INKLSQTSHLVLPASHADGTQLTQVASFATPDKKTAAIAYTSRLGNGKPSRLDIDOK 213
Db      165 DTEKWFELSH-----DQKNVYSIYA--NQKTSYCWMLKQYFNKNNTYDHLNVSIN 211
QY      214 EIIDGEELF--NAVQTLKLTIPNGYKSGIDQDAFVNDK-----NIAEV- 253
Db      212 RLTEAEFYAFDDFSQTIKLT-MNSYQTVNIDVNFNNLCLALRLPLSLERFNILNIR 270
QY      254 -----NLPELETISDYAFAMSLKQVXLPDNLKVYIGELAFPDNQIGK----- 297
Db      271 SSYTRNQYNFEKIGELLETFEAVVFSHRHLOGIHLOVPCFAFOYLWNSSSQISVKDSQLO 330
QY      298 LY-----LPRHLIKLAERAFKSNRIQTVEFLGSKLKVIGEAS 334

```



```

Db      331 VYSESTDLKLVDTNKVQDYFKELQEPFR-LTHVSQALPVSAATNAVENLNLKKVKAH- 388
QY      335 FQNNLNRLVMLPDGLEKIESEAFGNPGEDEHYNNQVLRTRTQGNPHOLATENTYVNDK 394
Db      389 ----NLNLVSI-----TQNFDFYFVNLQHLKLEFLEPNILTKQKL-----E 428
QY      395 SLWRATPDMYTKWLEEDF-TV--QKNSVTGFSNKG--LQKVRNKNLEIPKQNGITIT 449
Db      429 NLLLSIKQSKNLKFLRLNFYTYVAQETSRKQILKQATIKNLKNNKQOETPETKDETPS 488
QY      450 EIGDNAFRNVDFQSKTLRKXYDLBEIKLPST--IRKIGAFQSNLKSFEAS--EDLEE 504
Db      489 E-STSGMKFEDHLESELEEDF-SVNLQATQETVDSLHKLLIRSTNLKKFKLSYKEMEK 546
QY      505 IKGAF--MNNRIGTL-DLKDKLIKIGDAFHNHIAVLPEVS-----QEI--G 550
Db      547 SKMDTFIDLKNIYETLNNLKRCVSNISNP--HGNISYELTNKDSFYYKFKLTINQELQHA 604
QY      551 RSAFRONGALHLMFIGNKYKIGEMAFLSNKLSEV-----NLSEQKQKLTIEVQAFS 602
Db      605 KYTFKQK-----EFQENNVKS--AKIESSSLESLIEDIDSLCKSIASCQNLOVY----- 650
QY      603 DNALSEVLPNLTQITREAFKRNHL--KEVKGSSSTLSQITFNA-FDQNDGKRFEGKV 658
Db      651 -NITASLIPNNIQ--KNPFKNPLLFQKQFQELKNLENVINCILDQ----- 695
QY      659 VVRTHNNSHMLADGERFIIDPDKLSSTMVWLEKVLKIEGLDYSLRQTTQ----- 709
Db      696 -----HILNSISEFLKNNKIKAFILKRYVLYLQY--LDYTKLFTLQQLPELNQVY 745
QY      710 --TQREMTAGKALLSKSNLRQGEKQK-----FLQEAQFLLGRVLDKALA--KA 756
Db      746 INQLEELTVSE---VHKQWENHKQKAFYEPLCEFIKESQTLQIDFDQNTVSDSI 801
QY      757 EKALVTKATKNGHLERSINKAVLAYNNSAIKKANVRKLEKELDLTDLVEGKPLAQ 816
Db      802 KKLIESISESKYHHYLRNLPSQ-----SSSLIKSEN-----EIQELLKACDEKGVLYKA 851
QY      817 TMVQGVYLLKTPPLPEYTYIGLVNVPD 843
Db      852 -----YKFPPLCLP-----TGTYVD 866

```

```

RESULT 13
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugenii
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; PRIOR FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

```

```

Query Match      3.6%; Score 192.5; DB 4; Length 2662;
Best Local Similarity 20.0%; Pred. No. 0.0002;
Matches 242; Conservative 170; Mismatches 433; Indels 365; Gaps 59;

```

```

QY      21 YSQEVYGLEREESVQEQTQASASEDDWFEEDN--ERKTNVSKENSTVDTEVS----- 70
Db      507 YDNLVLDYEQLTRKEKEMELKLKEKNLDDEFEALERTKDKDQEMQILHETISNLKNLVKHR 566
QY      71 DLFSQNSNNSSKTESVSDPKQVPKAKREV--TQASNSNDAS---KVEVPKQDTAS 125
Db      567 EYVNOOLENELSSKVELLRKEDQIKLQYIDSOQLENIKMDLSYSLESIEDPKQ---- 622
QY      126 KKETL--ETSTWEAK--DFVTRGDTLVGFSKSGINKLSQTSHLVLPASHADGTQLTQVA 180
Db      623 MKQTLFDAETVALDAKRESAFLRSENLE--LKEKMKELATTY-----KOMENDIQLYQ-- 673
QY      181 SFAFTPDKKAIAEYTSRLGENGKPSRLDIDQKEIID-EGEIFNAY-QLTKLTIPNGYKS 238
Db      674 -----SQLEAKKQVVDLEKELQSAFNEITKLT----- 701
QY      239 IGQDAFVDNKNIAEAVNLPESELETISDYAFAMSLKQVKLPDNLKVIGELAFEDNQIGKL 298
Db      702 --SLIDGX-----VPKDLLC-----NLELEGKITDLOKELNKEV 733
QY      299 ---YLPRHLIKLAERAFKSNRIQVE-----FLGSKLVIGFASFDNNLRN 342
Db      734 EENALREEVITLSELKSLPSEVERLRKEIQDKSEELHITSEKDKLFSEVHAKESRVQG 793
QY      343 VMLPDGLEKIESFAFTGN--PGDEHYNNQVLRTRTQGNPHOLATENTYVNP----- 392
Db      794 LLEIGKTKDGLATQSNYKSTQDEFQNFETLHMDFEQKYKMYLEENENKQEIIVNLSKE 853
QY      393 ---DKSLWRATPDMY-----TKWLEEDFTYQKNSVTGFSNKG--LQKVRNKNLE 438
Db      854 AQKFDSSLGALKTELSYKTOELQEKTRVQERLNEMEQLKEQLENRDSPLQTVREKTLI 913
QY      439 IPKQNGI---TIT-----EIGDNAFRNVDFQSKTLRKXYDL 472
Db      914 TEKLQOTLEEVYTVLTQEKDLDKQLQESLQIERDQLKSDIHDTVMNMNIDTQEQ--RNAL 971
QY      473 EIKLPSTIRKIGAFQSNLKSFEASEDLEIKGAFMNNRIGTL--DLKDKLIKIGD- 529
Db      972 SLKHQETI-----NTLK--SKISEEVSRLHMEENTGETKDFQKQVAGIDKK 1017
QY      530 ---AAFHNHIAIVLPESVQEIQRSAFRONGALHLMFIGNKYKIGEMAFLSNKLSESVN 586
Db      1018 QDLKAKNTQTLTADVKNDEIIEQQRKIF-----SLIQEKNELQOMLESV- 1061
QY      587 LSEQKQKLT-----LEVQAFSDNALSEVLPNLTQIRE-EAFKRNHL--KEVKGSSSTLS 638
Db      1062 IAEKEQLKTLKENIEMTIENQELR--LLGDELKQOETVAQEKNAIKKEGELSRTCD 1119
QY      639 QITFNAFDQNDGDKRFGKVVVRTHNNSHMLADGERFIIDPDKLSSTMVWLEKVLKIEG 698
Db      1120 RLA-----EVEEKLKESQOLOEQOQLN--VQEMSEMOKKINEIEN 1161
QY      699 LDYSTLRQTTQTPREMTAGKALLSKSNLRQGE-----KQKFLQEAQ--FPLGRVLD 750
Db      1162 LKNEL--KKNELLHEMETERLELAQKLNENYEVKSIKERKVKLELOKSPETERDHLR 1219
QY      751 KAIKAE-KALVTKATKNGHL-----LERSIN-KAVLAYNNSAIKKA----- 791
Db      1220 GYIREIATGLQTKELKIAHILHKEHOETIDELRRSVSEKTAQIINTQDLEKSHTKLQE 1279
QY      792 -----NVKRLK-----ELDLTDLVEGKGPLAQATWQGVYLLKTPPLP 832
Db      1280 EIPVLHEQELLPNVKVSETQETWNELELLTE-----QSTTKDSTTLARIEME-- 1328
QY      833 EYVIGLVNVPKSGKLIYAL-----DMSDTIGEG-----QKDAYGN 868
Db      1329 --RLRLNEKQESQEBIKSLTKERDNLKTIKEALBVKHDQKKEHIRETLAKIQESQSKQE 1386
QY      869 PILNVD--DNEGHTLAVATLADVEGLYIKDILNSSLDKIKATROIPLAKYHRL-GIFQA 926
Db      1387 QSLNMKEKQNE-----TTKIVSEWQFQPKPD--SALLRIE-IEMLGISK--RLQESHDE 1435

```

QY 927 IRNAAEADRLLPKTPKGYLNEVPNRYKKOMEKOLKPVDTKPIFNKALPNEKVDGDRRA 986
Db 1436 MKSVAKKEDL-----QRLQEVLOSSESQDLKENIK-----EIVAKHLETE-----EELK 1479
QY 987 KGH---NINAETNNSVAVTPIRSE-----QOLHSQSDVNLPGT 1022
Db 1480 VAHCCLKEQETINELRNLSEKETISTIQKLEAINDKLQNKIQEIVEKEBQNLKQI 1539
QY 1023 SSKNFIYEI 1032
Db 1540 SEVQENVNEL 1549

RESULT 14
US-09-107-532A-3959
Sequence 3959, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 3959:
SEQUENCE CHARACTERISTICS:
LENGTH: 1221 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1221
SEQUENCE DESCRIPTION: SEQ ID NO: 3959:
US-09-107-532A-3959

Query Match 3.5%; Score 189; DB 4; Length 1221;
Best Local Similarity 19.0%; Pred. No. 0.00012;
Matches 228; Conservative 174; Mismatches 403; Indels 394; Gaps 61;

QY 4 KHLKTLAL-ITVSVTVYSQEVY---GLERESVKQEQOTQASSEDWFEEDNERKTNS 59
Db 14 KKLKLSIALPSLITMSTARQISVAGTSTKAVVEESGDKVAIDKYNKNEAKATVQ 73
QY 60 KENS---TVEETVSDLFSDGNS--NNSSKTESV---VSDPKQ--VPAKPEVTQEASN 108

Db 74 GADGIMYVVVDNDPEKLLKASPLKNQOENTGAVEVGDEKQEKADKANAKVKNFDDI 133
QY 109 SSNDASKVEV---PK-----QDTASKKETLETSTWEAKDFV--TRGDTLVGFSKSGINK 157
Db 134 DYNDISSLOSALNDPKFPFLSEDELNEKLQTKNTFESEALMKTR-----GYFSQRILE 188
QY 158 LSGTSHLV-----LPSHAADGTQLTQVASFAPTPDKKTAIAEYTSRLGENG-----KPS 206
Db 189 LSNLNDVPRKAKVSRHALVSNRASQVSETRINPNPGYPIVYILIR--SNGYSYTENYAI 246
QY 207 RLDDQKEII-----DEGEIFNAYQTLTIPNGYKSG----- 240
Db 247 RRTIDKEAFCLQAGVQATNTYTGTSGYLSNSQREMISSILSAYKELGWMANDAGKOR 306
QY 241 ---QDAFVDNKNIAEVNLPESLETISDY-----APAHMSLKQVKLPDNLKYI 284
Db 307 WIATQMLWENSN---EKPKSITVTPNWKNGATKWKNDLVARANAIRVTRKWDSTSKTY 362
QY 285 ---GELAF-----DNQI-----GKLYLPRHLIKLAERAFKSNRIQVEFL----- 323
Db 363 DEGDKATFKVTNTYGDHNLHVKSVSQGGKATISNGTLLVDTTGATGKI--TVDMIKGSH 421
QY 324 ---GSKLVIGEASFQDNNLR--NVMLEPDLE--KIESE-----AF 357
Db 422 SDAHPTVWSNGSYQKLVGTSFYISTSVTVNVRKRGVEVLKVDSETKPLSGATFRFSY 481
QY 358 TGNPGDEHYNNQVLRTRTGQNPQOLATE---NTYVNPDKSLMRATPMDYTKLEDE 413
Db 482 SGTKD-----VTTDSNGKAKLTERLKAGTVK-----VKEIKAPN 517
QY 414 TYQKN---SVTGSNKGLOKVRNKNLEIPKQNGITITEIGD---NAFRNVD----- 461
Db 518 GYQDLSSEFSITVKENQNTTTRTNK---KSTGSVEIEKIGDLGLPNVYFTIYNS 572
QY 462 QSKTLR-----KYDLEIKLPS--TIRKIGAFAPQSNMLK--SFEASDLEIEKEGA 509
Db 573 DNKVVKNLKTDSNGKIKVDLOFGKYAAVEKQGVTVGPDGKKYNFELISQDTPFTK--- 629
QY 510 FMNRIGTLDLKDILIKIGDAFHHNHIVAYLVPESVQEIGRSAFRONGALHMFIGNKV 569
Db 630 ---PAKVKTNIYEVSISEKVDNMSQVLKN----- 659
QY 570 KITGEMAFLSNKLSEVNL--SEQKQKLTIEVQAFSDNALSEVLPNLTQIREAFKRNH 627
Db 660 WSADPNLYQYDITGNLKKSGERNVQSFIEGFYDN-----KANVDSKSEVY--- 707
QY 628 LKEVKSSTLSQITFNAFDQNDGDKRFGKVVVVRTHNNSHMLADGERFIIDPKLSTMV 687
Db 708 ---VGSTNVT---NSFDINNDT-----NNGKVIKAKTSVLTNNDFYKSY 747
QY 688 DLEKVLKIEGLDYSLRQTTQTOFREWTTAGKALLSKSNLRQGEKQFLQEAQFLGRV 747
Db 748 NLRITMKI-----KKSSTLENTERQKNL----- 769
QY 748 DLDKAIKAEKALVTKKATKNGHLERSINKAVLAYNNSAIKKANVRRL-EKELDLTDL 806
Db 770 ---TIVKNSKVTVGSKISKSA---SSNDVETVLYS---RKVTINHIDEKDKHLKQD 817
QY 807 VEGK--GPIAQATMVQGVYLLKTPLPBEYYIGLNVPFDKSGKLIYALDMSDTIGEGQKD 864
Db 818 IDKYDGETYE-----YKPRITDL-----FDKDGNNY---KSDVIHKGKID 854
QY 865 AYGNPILNVEDNEGYHTLAVATLADYEGLYIKDILNSSLDKIKARQIPLAKYHRLGIF 924
Db 855 ---GKDIYL---NTPYHI-----PVLNVNDRI-----QIDTAR----- 882
QY 925 QAIRNAAEADRLLPKTPKGYL-NEVPNRYKKOMEKOLKPVDTKPIFNKALPNEKVDGD 983
Db 883 -AVKN-----GYLPFKIELSKKEEYEKELEKIVFAYKITD--IDNMKVAVYD 925
QY 984 RAAKGHNINAETNNSVAVTPIRSEQOLHSQSDVNLPGTSSKNFIYE-----ILGYVS 1037

